

A true test of leadership

Muammar Gaddafi, the Libyan leader, can demonstrate his aspirations to statesmanship by using the occasion of the XV International AIDS Conference to free health workers unjustly sentenced to death for infecting patients with HIV.

Valya Cherveniyashka; Snezhana Dimitrova; Nasya Nenova; Valentina Siropulo; Kristiana Valcheva. These five Bulgarian nurses, along with Ashraf Ahmad Jum'a, a Palestinian doctor, were sentenced to death by firing squad by a Libyan court on 6 May. Their crime: deliberately infecting 400 children with HIV at the al-Fateh Hospital in Benghazi in 1998, causing the death of 40.

The convicted health workers, who have languished in prison since 1999, will appeal. They need our support. Amnesty International and other human-rights groups have already denounced the convictions, arguing that confessions were extracted by torture.

The convictions also run counter to the scientific evidence. Luc Montagnier, whose group at the Pasteur Institute in Paris discovered HIV, and Vittorio Colizzi, an AIDS researcher at Rome's Tor Vergata University, were asked by the Libyan government to examine the scientific evidence. Their report, submitted in April, should have removed any suggestion that the accused were guilty.

Montagnier and Colizzi's genetic analysis of viruses from the infected children indicated that many of them were infected long before the Bulgarian nurses set foot in Libya in March 1998; some of the infections probably dated as far back as 1994. The children also showed a high incidence of co-infection with hepatitis B and C, suggesting that the underlying problem was an infection from a common source, spread by poor hospital hygiene. This is a tragedy, but such problems aren't unique to Libya.

The analyses also showed that the infections were caused by subtypes of A/G HIV-1 — a recombinant strain common in central and west Africa, and known to be highly infectious. The samples showed little variation, again suggesting a common origin, probably spread from an already-infected patient via dirty catheters and syringes.

Another nurse, who reported accidentally injecting herself while handling blood samples, carried an almost identical virus.

Montagnier argues that, in a further travesty of justice, the judgment was based at least partly on a mistranslation from English to Arabic of the term 'recombinant'. This term, used by Montagnier and Colizzi to mean the natural recombination of wild viruses, was translated in the judgment to mean a 'genetically modified' virus, says Montagnier, implying human manipulation.

The verdict slaked the mob's thirst for vengeance, being greeted with rejoicing in Benghazi. But it should chill the blood of anyone who cares for justice and the use of scientific evidence in its name.

In a letter earlier this month, Montagnier and Colizzi, backed by leading virologists and physicians, asked Libyan leader Muammar Gaddafi to commute the death sentences, dismiss the cases and review how the trial was conducted. Montagnier wrote a second letter to Gaddafi on 6 July, arguing that clemency "will certainly magnify the good image of Libya in the world".

Further pressure is warranted. After decades as a pariah, Libya is now taking steps to rejoin the international community. Gaddafi helped whip up domestic feeling against the health workers, arguing in 2001 that their actions were part of a plot orchestrated by the CIA and Israel's Mossad. By recognizing the injustice that has since been done, he could send a clear signal that he wants to put his past behind him.

Opening the XV International AIDS Conference in Bangkok, Thailand, on 11 July, United Nations secretary-general Kofi Annan argued that defeating HIV requires leadership at the highest level. Gaddafi has a chance to show the world that he now understands that true leadership means embracing justice, compassion and a respect for scientific evidence. ■

Science on show in Stockholm

Introducing the first Europe-wide meeting for scientists and science's stakeholders.

Why do science's messages so often get lost in the midst of media coverage of controversies such as genetically modified crops? If they didn't, would a regulatory process that involved more public participation help or hinder technological development? How can we separate hype from reality in fields such as stem-cell research and robotics? What are the neural underpinnings of our awareness of action? And, by the way, how can light be brought to a standstill?

The chances are that, as a *Nature* reader, you are interested in all of these questions. They and many others are being addressed in Stockholm, Sweden, on 25–28 August, at a meeting for scientists, policy-makers, the media and the wider public: the EuroScience Open Forum 2004. For the conference programme, see www.esof2004.org.

Nature should declare its interest, having supported the forum since its conception (see *Nature* **423**, 571; 2003). We are organizing two sessions: one on the prospects and challenges presented at the European level by infectious diseases; the other examining the predictability or otherwise of extreme climatic events in Europe. We are

also organizing talks and workshops on careers themes, including one of universal interest, on how to negotiate your salary.

There are obvious parallels with the annual meeting of the American Association for the Advancement of Science (AAAS). In a cynical age, some observers have questioned the need for such large gatherings. And there are bound to be grumbles about holding the inaugural European meeting of this type at the tail-end of the holiday season. But if nothing else, the AAAS meeting provides a valuable focus for mainstream media treatment of issues of scientific importance. If the EuroScience forum can achieve a reasonable fraction of this coverage, and a measure of public participation, it will be a welcome addition to the annual meetings calendar.

As its name suggests, the meeting includes a specifically European agenda, with sessions that address both professional issues (such as the commercialization of Europe's universities) and topics of broader topicality (for example, the continent's transformation to a low-carbon economy). But science is global, and so are most of the topics that will be discussed in Stockholm. We hope to see you there. ■





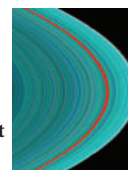
Money plan
Britain backs
funding boost
for science
p280



Out in the cold
Bush administration
criticized for
rejecting Nobellist
p281



Party lines
German reforms
stall as political
in-fighting breaks out
p283



Mucky photos
Grubby state of
Saturn's rings
revealed
p285

Aid agencies predict victory for HIV unless cash crisis is solved

Erika Check, Bangkok

For the first time since HIV was identified two decades ago, the global effort to contain its spread is beginning to bear fruit. But this progress could be derailed by lack of funds and disagreements over the best way to prevent transmission, speakers told the XV International AIDS Conference in Bangkok, Thailand, this week.

Since 2000, global spending on AIDS in developing countries has grown by US\$4.8 billion to \$6 billion, according to a report released on 12 July by the Joint United Nations Programme on HIV/AIDS (UNAIDS). And the investment is starting to pay off, leaders at the Bangkok meeting said.

"Finally the political, technical and financial resources are starting to move, and that changes everything," said Peter Piot, executive director of UNAIDS. "Now, we must work even harder to make that money work on a massive scale."

One of the main instruments for the effort, the Global Fund to Fight AIDS, Tuberculosis and Malaria, had barely opened for business when the last international AIDS conference was held in Barcelona in 2002. But Richard Feachem, its director, told the meeting that its efforts have already reached some 2.3 million people. The fund has distributed antiretroviral drugs to 5,200 HIV-positive patients, and has cured 45,000 people of tuberculosis.

Yet many large donor nations still prefer to invest in their own, bilateral AIDS programmes, and the fund is struggling to meet its income targets. It faces a cash shortfall of \$100 million this year, and a huge \$2.6-billion shortfall on a projected budget of \$3.5 billion next year. Overall, UNAIDS says that worldwide spending on AIDS in 2005 will fall \$6 billion short of the \$12 billion needed to fight the epidemic effectively.

Donor nations have scrutinized the Global Fund's plans intensely, its officials say, partly because it is the first major aid effort to give money to nations that have written their own action plans, rather than exporting pre-made programmes overseas. Feachem said the fund has proved that this model works, and speakers from developing countries agreed that it meets their needs better than the old model.



In your face: AIDS activists deface a picture of Tony Blair in a Bangkok protest against the G8 nations.

"Many donors are driven by their national ideologies, rather than by scientifically proven approaches and responses," says Bizwick Mwale, executive director of the National AIDS Commission of Malawi.

As a result, aid recipients say, programmes are often out of tune with local needs. Donors tell local officials what to say, says Serara Mogwe, of the University of Botswana School of Nursing in Gaborone, "and we end up talking to our people in a strange language that they don't understand". In the past, Mogwe adds, the emphasis has been on sex education and condom use. But, she says, the public discussion of either is almost impossible in Botswana, which has the second-highest rate of HIV infection in the world.

Three-point plan

Now, the United States — the largest donor for AIDS prevention and treatment — is promoting a mantra known as ABC: abstinence, be faithful and use condoms. This approach was widely castigated in Bangkok, where 17,000 scientists, activists and officials have gathered for the AIDS meeting. Activists and some researchers are particularly critical of a congressional stipulation that requires one-third of the money allocated to preven-

tion programmes under the US President's Emergency Plan for AIDS Relief to be used for projects in abstinence and monogamy.

"You're not doing what countries want or what people with AIDS want," Gregg Gonslaves of the US activists' group Gay Men's Health Crisis told a US official at a panel on 12 July. "You're trying to please George Bush's conservative base."

But Randall Tobias, global AIDS coordinator at the US Department of State, said Bush has not turned his back on condoms. "Condom use is appropriate when people decide to engage in high-risk behaviour," said Tobias. "Condoms are an important part of our overall strategy."

Bush was not the only world leader drawing fire at the conference. In a demonstration on 12 July, protesters branded leaders of all the G8 club of rich nations 'AIDS accomplices' for not giving enough to the Global Fund and other AIDS efforts.

Feachem, meanwhile, warns that the fund will have to postpone its next round of grants until 2007, unless it gets more assurances of support soon. "The consequences will be catastrophic," he says. "We will not win the fight against AIDS."

♦ www.aids2004.org

Britain spends to secure scientific growth

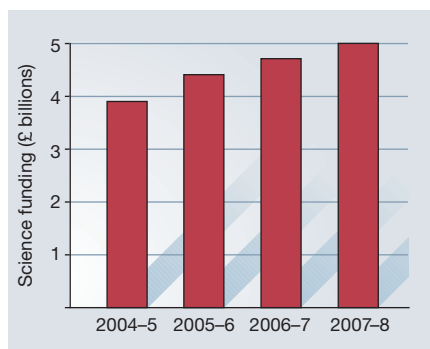
Jim Giles, London

UK researchers will find more money in the coffers of those funding them over the next three years, thanks to spending plans announced by the government on 12 July.

The plans call for the country's two main science-funding streams, which cater for basic research grants and university running costs, to be boosted by 5.8% a year in real terms between 2004–05 and 2007–08. That does not quite match the 10% annual increases called for in the previous 2002 spending review (see *Nature* 418, 261; 2002) and seen over the past few years. But it will still take government funding for science and engineering to £5 billion (US\$9.3 billion) annually, up from £3.9 billion in 2004–05 (see graph).

The figures, unveiled by the chancellor of the exchequer, Gordon Brown, as part of a ten-year science strategy, were accompanied by a slew of smaller measures that will also find favour among researchers. An extra £80 million over three years has been allocated to help fund salaries associated with basic research grants, for example, easing complaints that these costs were being ignored by funders.

"There is much in this that appears to be genuinely good news for the science community," says Peter Cotgreave, director of the pressure group Save British Science. More good tidings were supplied by the Wellcome Trust. To coincide with the chancellor's speech, the medical charity has announced that it will invest £1.5 billion (US\$2.8 billion) in Britain



Hand out: Gordon Brown announces annual increases in funds for basic grants and university upkeep.

over the next five years, which will maintain its spending at current levels.

In the long term, Brown says he also hopes to raise combined government and business research and development funding from the current value of 1.9% of gross domestic product to 2.5% by 2014. This ambitious target would bring Britain in line with competitors such as the United States and Germany — which stand at 2.7% and 2.5%, respectively. If this happens, and funding continues to rise at 5.8% over the next decade, the level of basic research grants may double as hoped (see *Nature* 429, 688; 2004). But Brown did not spell out which science-funding streams would be favoured under any future increases.

Critics question whether the strategy contains the new ideas needed to achieve this goal. Science-policy experts say that reaching



the 2.5% target depends on increased private-sector funding. America's top 700 companies reinvest around 5% of money from sales into research and development, for example, but UK companies spend only half of that.

Yet the strategy focuses instead on boosting existing measures, such as grants to help universities work with industry. "There are no new fiscal measures and that is disappointing" says Cotgreave.

Lack of action on researchers' salaries was the other notable omission from the strategy. PhD stipends will remain frozen in real terms, although they have grown substantially over the past three years. Greater concern centres on pay for academics. It is unclear whether increases to the university-funding stream will be channelled away from infrastructure projects and into improved wages. ■

Authors urged to come clean on competing interests

Jonathan Knight, San Francisco

Many authors continue to keep financial conflicts of interest to themselves, despite ever stricter journal policies requiring full disclosure in published articles. That is the conclusion the Center for Science in the Public Interest (CSPI), a Washington-based pressure group, has reached after reviewing all the articles in four major medical research journals over a three-month period.

The problem is the result of a combination of lax enforcement by editors and loopholes in disclosure policies, says Merrill Goozner, a CSPI researcher and the author of the review. "Our study is a reminder that journals need to be vigilant," he says.

Concerns that researchers' financial interests might influence their objectivity or methods have been increasing for 20 years as ties between industry and the academic sphere have broadened. Yet research journals have been slow to respond. By 1997, only 16% of the 1,400 top biomedical journals had any competing-interest

disclosure policy in place, and enforcement was patchy at best (see *Nature* 411, 3; 2001).

Goozner examined articles published from December 2003 to February 2004 in the *New England Journal of Medicine*, the *Journal of the American Medical Association*, *Environmental Health Perspectives* and *Toxicology and Applied Pharmacology*. He used websites and public databases to investigate the lead authors of each article in which a declaration of no competing interests was made. In a report released on 12 July, Goozner listed 13 studies, out of 163, that failed to comply with the publishing journal's disclosure policy.

In one example, an author applied for a patent on a reagent that his paper suggested would be valuable in cancer research. In another, the authors on a paper about coronary heart disease were paid consultants to more than 20 companies in the heart-disease field, but did not disclose this. According to the CSPI report, the authors confirmed the relationships by

e-mail, but said "None of them had anything to do with the paper. Why give them free advertising?"

Goozner says it should be up to the readers to decide. Because journals ask authors to declare only 'relevant' or 'directly relevant' ties, authors can too easily rationalize their omissions, he claims. Goozner's solution is to require disclosure of all ties to private firms and all patents regardless of relevance, and then let the editors decide which to publish.

The situation may be improving. Although it examined different journals, a 2001 study found that only 1% of papers disclosed a conflict (*Sci. Eng. Ethics* 7, 205–218; 2001). In the current study, that rate was up to 20% of papers, with those failing to declare an interest making up only 8%. That at least brings the problem into a manageable realm, says bioethicist Mildred Cho of Stanford University, California. "We should be able to deal with this in the future." ■



Torsten Wiesel: the Nobel laureate was 'too critical' of the US president to get a place on a scientific panel.

Bush accused of trying to foist favourites on health agency

Emma Maris, Washington

The former head of a US health-research agency has accused the Bush administration of rejecting most of his nominations for his centre's advisory committee — including a Nobel laureate — on political grounds.

Gerald Keusch was director of the Fogarty International Center, the branch of the National Institutes of Health that supports projects in developing countries, until last December. In a report released on 8 July by the Union of Concerned Scientists (UCS), he charges that the process for appointing members to the centre's advisory board was subject to excessive interference. The board is involved in, among other things, approving Fogarty grant applications and advising on the focus of the centre's research.

Out of 26 people proposed by Keusch for board membership during the Bush years, 19 were rejected by the office of Tommy Thompson, the secretary for the Department of Health and Human Services, Keusch says. Along with letters rejecting his candidates, he received resumés for other people that Keusch described in an interview as "lightweights" with "no scientific credibility". He adds that he felt under subtle pressure to nominate them for membership of the board. "I was never told to take any of them," he says, "but the hint was clear."

Keusch's first round of nominees to Thompson, submitted in 2001, included Torsten Wiesel, who won the 1981 Nobel Prize in Medicine for his work on visual processing in the brain. When Wiesel's name was rejected, Keusch climbed the bureaucratic ladder to find out why. Eventually, an official in Thompson's office told him that Wiesel had "signed too many full-page letters in *The New York Times* critical of President Bush", according to the UCS report. Keusch says

that other candidates were rejected because they were associated with reproductive health or abortion rights. Last December, Keusch resigned and took a position as associate dean for global health at the Boston University Medical Campus. "I felt that I had ceased to be effective," he says.

Bill Pierce, a spokesman for the health department, points out that under the law the health secretary appoints the board. The secretary takes recommendations from other sources as well as the director of the Fogarty, Pierce says, adding that not only is scientific expertise considered, but also the mix of gender, race, geography and political opinion. "We want diversity of thinking," Pierce says.

The 34-page UCS report documented several instances of what it described as political meddling in science by the Bush administration. The report follows a similar publication that was released in February with a statement signed by a slew of scientific luminaries (see *Nature* 428, 250–251; 2004).

The latest report includes allegations that the Department of the Interior and other federal agencies suppressed the results of research on the environmental impact of a controversial mining practice called mountaintop-removal strip mining. It also says that potential appointees for several bodies were asked whether they liked US President Bush or had voted for him.

Pierce counters that the UCS itself is politically biased, noting that several of its prominent members, including chairman Kurt Gottfried, a physicist at Cornell University in New York state, have contributed money to Democrat causes and candidates. "To have an opinion is fine," he says, "but to pass yourself off as independent when you have an agenda is not. They are the ones who are politicizing science."

Indian scientists welcome broad increase in funding

K. S. Jayaraman, New Delhi

Science spending is to get a major boost under a budget laid out on 8 July by India's new Congress-led government.

Total expenditure on research and development will increase by almost a quarter to Rs152 billion (US\$3.3 billion) under the budget for the financial year which began in April.

"There will be a 15% hike in funding for all our major projects — and that isn't bad," says Valangiman Ramamurthi, secretary of the Department of Science and Technology.

Indian scientists are delighted with the budget. "I am very pleased," says Raghunath Mashelkar, head of the Council of Scientific and Industrial Research, the largest science agency in India.

The increase is widely distributed across different spheres. Funding for fusion research at the Department of Atomic Energy will double last year's budget to Rs700 million, for example, and funding for multidisciplinary research projects at the science department will grow by Rs400 million to Rs2.4 billion.

The budget allocates almost Rs5 billion — one fifth of the entire space budget — to the development of a new rocket, the Geosynchronous Satellite Launch Vehicle Mk-III, giving it the largest funding of any single science or technology project in the country. Gopalan Madhavan Nair, the space secretary, says the rocket will be able to launch satellites weighing up to 4 tonnes and will fly by 2007.

A 21% increase in support for space research will also allow scientists to begin work on a Rs3.5-billion lunar orbiter mission planned for 2008 and on a recoverable satellite that could be a forerunner for a manned space mission.

The national budget — which takes advantage of strong economic growth to boost spending, despite India's growing budget deficit — includes billions of dollars for rural development. Some of the science and technology money will be channelled towards this. There will be more support for agricultural biotechnology, for example, and for technology to clean water supplies. This will include funding for the country's first commercial-scale desalination plant, in Chennai, which will be followed by a number of others.

For more news and analysis go to
news@nature.com
www.nature.com/news

Institute doomed by loss of interest in basics

David Cyranoski, Tokyo

Eighteen years ago, it was a shining example of industry-sponsored science in Japan. But the Biomolecular Engineering Research Institute (BERI) in Osaka is to close next year, despite its scientific success.

The institute's plight reflects the wider fate of industry-sponsored science in Japan over the past 20 years. The once bountiful well of private-sector support for basic research has dried up, as grand corporate ambitions have shrunk in more austere and uncertain times.

Founded in 1986 as the Protein Engineering Research Institute, it was funded by the industry ministry and 18 companies. Its mission was to develop basic science that would underpin drug development. It now has an annual budget of ¥1.1 billion (US\$10 million) and some 70 researchers.

Over the years BERI, as it became in 1995, has notched up important findings, particularly in the analysis of protein structures such as the glutamate receptor, which plays an important role in the nervous system (N. Kunishima *et al.* *Nature* **407**, 971–977; 2000). “BERI researchers put out outstanding papers one after the other,” says Shigeyuki

Yokoyama, director of the protein group at the Genomic Sciences Center in Yokohama. “They really led Japan in the early days of protein structure analysis.”

But in April, the institute's board of trustees, which includes representatives of the remaining nine companies that provide 30% of BERI's funding, decided to drop their support. Chairman Tadashi Hirata, chief executive of drug company Kyowa Hakko, refused to comment on the reasons for the decision.

“BERI has been very productive,” says Mitsuru Takeshita, head of the biotechnology division in the economic ministry's New Energy and Industrial Technology Development Organization, which provides the other 70%. “The problem is not whether BERI has been putting out results — which it has. Rather it is an internal organizational problem. The closing is very regrettable.”

The institute's director, Kosuke Morikawa, says that BERI is closing because the gap between scientific discovery and industrial application is too large for sponsors to justify the cost. “From a research perspective we have been extremely successful and we have also trained many of the

leaders of industrial science,” says Morikawa. “But the companies have not retained much interest.”

BERI's method of studying proteins one-by-one is also out of step with the move towards industrial-scale approaches such as the Protein 3000 project, Japan's effort to solve the structure of 3,000 molecules. “Groups around the world are finding that a bigger scale is better,” says Yokoyama, whose group is leading the project.

Morikawa says he has courted patrons who might be able to support the institute, but so far has had no success. When he asked Hirata what to do, he says he was told to relaunch BERI as a small business. “I don't have the connections or knowledge to do that,” he points out.

BERI's researchers may struggle to find employment at a time when Japan is trying to introduce more intense competition, with greater emphasis on industrial involvement in basic research (see *Nature* **429**, 207–221; 2004). Judging from the experience of BERI, which Morikawa envisioned as a model for such involvement, it will be a challenge for research institutions to meet these goals. ■

Flying labs aim to track pollution across the Atlantic

Amanda Haag, London

Scientists from North America and Europe took to the skies this week to track the movement of giant masses of polluted air.

During the six-week study, a tag team of five aircraft will follow parcels of air pollution as they travel from the east coast of the United States over the Atlantic towards Europe. An international team of some 100 scientists will pick apart the collected data.

Their mission: to understand the chemical interactions that take place as ozone-forming pollutants from factory smokestacks and car exhausts in the United States hitch a ride to Europe on prevailing winds — a journey that takes about four days.

The International Consortium for Atmospheric Research on Transport and Transformation (ICARTT) is launching the project to discover how atmospheric pollutants such as aerosols, free radicals and ozone behave chemically as they are transported from one continent to the next — and how they affect climate change and health issues.

The pollution that the researchers are



Plane truth: NASA's DC-8 forms part of a fact-finding fleet.

studying builds up over an area covering New York, Washington DC and Boston during periods of still weather throughout the year. When the weather changes, this mass of air moves from the lower part of the atmosphere to the upper atmosphere. Winds then push the air mass over to Europe. The lump tends to break up as it crosses the Atlantic, but scientists hope to track at least one parcel of pollution that holds together.

The aircraft will measure the size of aerosol particles and the chemical

constituents of the invisible cloud, using weather forecasts, models and measurements to stay on track.

Similarly equipped planes have been used in the past to study small sections of the journey, but no sustained mission has successfully tracked the pollution right across the Atlantic.

Satellites have been used to monitor air transport over the Atlantic, but cannot be used to pick up on the details of the chemistry.

The multimillion-dollar study may one day help to shape global atmospheric policies, according to the team behind the project.

Pollution isn't just a local problem, says its lead UK scientist, Alastair Lewis, an atmospheric chemist from the University of Leeds. Dust blows from Africa to North America, and Europe's smog may be caused in part by pollutants from the United States.

During the 2003 summer heatwave, there were about 2,000 more deaths than usual in Britain, some 21–38% of which were attributable to heightened atmospheric pollution (J. R. Stedman *Atmos. Environ.* **38**, 1087–1090; 2004). “Disentangling where it came from is very tricky,” says Lewis. ■



Caught in the crossfire: universities such as Leipzig may miss the chance to compete for extra money.

Political wrangling derails German university reforms

Quirin Schiermeier, Munich

An ambitious attempt to reward Germany's top universities looks set to fall victim to squabbling between the country's main political parties.

At a meeting on 5 July, the federal government and Germany's 16 *Länder* (states) failed to agree how to implement a competition to identify and reward the country's best universities. The €1.9-billion (US\$2.4-billion) programme is now on ice until November, when attempts will be made to renegotiate it. But after last week's unexpected failure, observers fear that this may never happen.

In Germany there is frequent wrangling between the federal government, controlled by a coalition of Social Democrats and Greens, and the states, mostly ruled by the Christian Democrats, which must approve most new laws. The power tussles thwart attempts at reform, and have been branded the 'German disease'.

Now science policy has caught the malaise. In February, Edelgard Bulmahn, the federal research minister, announced plans for a competition to start this summer. From 2006 to 2010, up to ten universities were to receive an extra €50 million each year for hiring top scientists, modernizing labs, and improving research and education (see *Nature* 427, 477; 2004). The effort was supposed to meet concerns that the German university budget is too thinly spread, and help a few elite institutions to compete internationally.

The *Länder* wanted to become involved, and agreed to contribute a quarter of the

money. But at last week's meeting, they made demands that could derail the idea. States controlled by Christian Democrats sought to delay the competition until their responsibilities for the universities have been clarified. A special commission is looking into the responsibilities of the federal and state governments, but it is unlikely to report soon (see *Nature* 423, 790; 2003).

"Bulmahn's wish to press ahead with the competition is politically motivated," says Peter Frankenberg, the science minister of the state of Baden-Württemberg in southern Germany. Like other Christian Democrats, he insists that there should be no time pressure to act.

In a 7 July statement, the 81 member institutions of the DFG, Germany's main grant-giving agency, called on policy-makers to settle their differences and get the competition moving. "Top university research in Germany needs extra funding to compete on a global stage," the statement said.

"There was a lot of enthusiasm here, and we reckoned on having a good chance of winning," says Franz Häuser, rector of the University of Leipzig. "Our best scientists have been busy for months preparing for the competition. Stopping the train at this point sends a highly disturbing political signal."

Now, Germany's cash-starved universities will have to wait for their bonuses. "It's the same old story," says Jürgen Mlynek, president of the Humboldt University in Berlin. "Everyone agrees on the content, but no one is prepared to agree politically." ■

Funding not keeping pace with science, warns Pasteur chief

Declan Butler, Paris

The Pasteur Institute — France's most prestigious private biology laboratory — is facing a financial crunch, warns its director-general, Philippe Kourilsky.

The institute, which employs about 2,500 staff studying microbiology and infectious diseases, says that public support for its researchers isn't keeping pace with its expanding activities. Over the past decade, says Kourilsky, income from contracts and patents has risen by about 40% to €45 million (US\$56 million) last year. Funds from donations and endowments rose by 70% to almost €60 million. The share of the budget met by public funding has plummeted to less than one-third.

"It's absolutely unfair," complains Kourilsky. "We do top science, and for every euro the government gives us, we generate two." The institute's scientific output far exceeds its 3.5% share of the French budget for life sciences, he says.

Kourilsky blames lagging public investment for the institute facing a €24-million shortfall in its €185-million budget for 2003. The institute expects to break even this year thanks to some large contracts — but Kourilsky says that he anticipates deficits in years to come.

The French government has strayed from a tacit agreement that public funding should cover about half of the institute's costs, he says. A science-ministry spokesman says that it is aware of the problem. He says that the government intends to provide "bonus funding" for some of France's top research centres next year.

During a previous financial crisis in 1965, Charles de Gaulle, then president of France, came to the institute's aid, saying that it was "as untouchable as the Eiffel Tower". But institute officials are less optimistic about being bailed out by the current president, Jacques Chirac. ■



Charles de Gaulle once likened the now-troubled Pasteur Institute to the Eiffel Tower.

Health agency gets cold comfort from House budget

Washington The US National Institutes of Health (NIH) has had its money worries confirmed by a key subcommittee in the House of Representatives. On 8 July, the labour, health and human services appropriations subcommittee recommended that the NIH receive a budget increase of 2.6% in 2005 — matching President Bush's request to Congress in February (see *Nature* 427, 475; 2004).

Research advocates say that they are deeply disappointed by the recommended \$727-million rise, which would bring the biomedical agency's budget to about \$28.5 billion.

"Investigators will be forced to make choices and perhaps cuts in personnel and equipment. Promising avenues of research will be abandoned," says Paul Kincade, an immunologist at the Oklahoma Medical Research Foundation in Oklahoma City and president of the Federation of American Societies for Experimental Biology. He calls the proposed increase "tragic and cruel".

Research advocates hope that a larger increase will be granted later this summer

by a corresponding Senate subcommittee. In that event, House and Senate negotiators will have to compromise.

Love in a time of pesticides for Asian rice farmers

Tokyo Scientists are turning their hand to writing soap operas in Vietnam and Laos. In an effort to curb the overuse of pesticides, two researchers from the International Rice Research Institute (IRRI) in Manila, the Philippines, are working with broadcasting companies in those countries to develop radio programmes that merge tips on the frugal use of pesticides with drama based on everyday tales of farming folk.

The researchers hope the show will correct some common misunderstandings and reduce pesticide use. "Our research shows that 80% of pesticides are spread at the wrong time for the wrong insect," says Kong Luen Heong, an IRRI entomologist who initiated the project. Farmers also don't understand that if insects eat rice leaves early on, the leaves will usually grow back, says Heong.

The show is based on *The Archers*, a UK programme with a similar agricultural mandate, says Heong. Sponsored by a US\$300,000 grant from the Rockefeller Foundation, Heong and his colleagues plan 104 programmes — two a week, starting this



Sound advice: a radio soap opera will tell farmers to avoid overusing pesticides.

month. "There will likely be stories of lovers and affairs," he says. "Some of the human conflicts will be tangled up in debates over the proper use of pesticides."

Environmental links to autism put to the test

London Can autism be caused by environmental factors — including immunizations and food? A British initiative is now set to tackle this question.

The Medical Research Council has granted £400,000 (US\$750,000) for a study into possible associations between autism and vaccine jabs, problems during birth, maternal and infant infections, maternal diet and fetal exposure to toxins.

The study will use data collected at the University of Bristol for a project called “Children of the 90s”, which is tracking some 14,000 children in an attempt to discover how the physical and social environment interacts with genetics to affect a child’s health, behaviour and development.

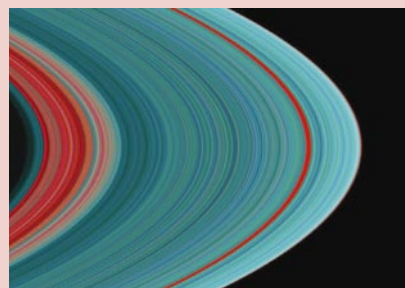
Kazakhstan launches plan for space programme

Moscow The central Asian nation of Kazakhstan plans to enter the space race, Prime Minister Daniyal Akhmetov announced on 9 July.

The country, which hosts the Baikonur launch site used by the Russian space agency, plans to spend US\$345 million on its own space programme. Russia will continue to use Baikonur and will collaborate in Kazakhstan’s plans, which include developing an environmentally friendly replacement for the Soyuz launch vehicle. Exhaust fumes from Soyuz, which ferries cargo and crew to the International Space Station, have been linked with environmental damage

Saturn’s rings yield their dirty secrets

London NASA’s Cassini spacecraft snapped this shot of Saturn’s rings as it flew through them at the beginning of July. The false-colour image captures ultraviolet light reflected by the rings, and highlights just how dirty they are. The ‘dirt’ — a mixture of mineral grains and organic molecules — shows up as red, with water ice appearing turquoise. The scale of the shot is impressive — the red band near the outer edge of the rings is 300 kilometres wide. This streak of dirt is made by the tiny moon Pan, which is



just 20 kilometres in diameter, as it circles Saturn, clearing a path through the ice and leaving a trail of minerals in its wake.

in Kazakhstan and neighbouring countries.

Kazakhstan also plans to send a cosmonaut into space in 2006 and to develop a satellite.

National Academy sets sights on satellite use

Washington The National Academy of Sciences has agreed to compile a wish list for Earth scientists keen on using satellites to study the planet. The Earth-science ten-year review, due to kick off at a workshop on 23–25 August in Woods Hole, Massachusetts, will be chaired by Richard Anthes, president of the University

Corporation for Atmospheric Research in Boulder, Colorado, and Berrien Moore, a climate modeller at the University of New Hampshire in Durham. The review panel, which plans to solicit ideas from the scientific community for new satellites or instruments to be piggy-backed on weather satellites, is expected to report by the end of 2005.

The academy has outlined ten-year priorities for several fields in the past, including astronomy, planetary exploration and solar–terrestrial physics. Such reports are used to help determine funding for projects at agencies such as NASA and the National Science Foundation.

Doing it for the kids

A few brave postdocs are mixing science with school teaching. Mark Peplow asks them what they give to, and get from, the children.

The defining moment of Vicky Taylor's postdoc came when she met an eight-year-old girl in a primary school. "I asked her, 'What do you think about science?' and she said straight out, 'Men are scientists,'" recalls Taylor. "But I'm a scientist," I said — and that totally changed her perception."

Taylor, an endocrinologist, is taking part in a scheme that pays researchers at Imperial College London to spend half their time teaching in local schools. INSPIRE (Innovative Scheme for Post-docs in Research and Education), now halfway through its trial period, is a response to fears that science teaching in the United Kingdom is in crisis.

The number of science students in UK schools and universities has fallen in most subjects over the past decade, eroding the supply of science teachers (see graphic). Some 800 new physics teachers were needed in 2000, but only 200 signed up for teacher training. Similar shortages and trends are seen in the United States, and when vacancies are filled, the teachers are less and less likely to have any science training.

INSPIRE is trying to reverse this trend. In collaboration with the pharmaceutical company GlaxoSmithKline, it offers three-year contracts to postdoctoral researchers, combined with a teacher-training course. Richard Sykes, former chairman of Glaxo-SmithKline and current rector of Imperial College, is the driving force behind the scheme, which began in September 2002. Since joining Imperial in 2001, he has become concerned by the shrinking pool of science applicants.

Sykes believes that INSPIRE can shape the next generation of scientists while they are still in the classroom. He had the contacts and clout to bring together drug-industry cash, academic expertise and government support to make it happen. Although INSPIRE's future funding is not yet secure, Sykes believes the scheme could be reproduced in universities and schools around the country.

Why encourage active scientists, not just graduates, to become teachers? Sykes's



Grass roots: letting in postdocs brings UK schools extra money for science staff and equipment.

response is blunt: "Most people who teach science haven't even been properly educated in it. As kids get older they become more sophisticated and ask more sophisticated questions — and if they don't get answers, they give up." Postdocs have the experience, background and depth of understanding to give them those answers, he claims.

Role models

John Leach, a professor of science education at the University of Leeds, supports this view. Many science teachers come straight from a degree that was largely about learning facts, he says, with little experience of doing science, either through a PhD course or industry.

All the INSPIRE postdocs agree that their experience allows them to act as role models, give careers advice, and talk about how science is used in the wider world. But there are some practical benefits too. Jenny Litten is a pig-nutrition researcher at Imperial College's campus in Wye, Kent, and is finishing her first year with INSPIRE. Sometimes the piglets in her research are stillborn, so she dissected one of these animals for a crowd of 60 pupils. The teacher she was working with had a physics background, and would never have done the dissection himself, explains Litten. "It's a different experience for the kids, because not many secondary schools demonstrate dissection any more," she says.

After three years, the INSPIRE postdocs can either continue their research or switch

to teaching. None of the original recruits has decided yet. But even if they stay in research, the schools they work in will still have benefited. Each school gets £40,000 (US\$75,000) from GlaxoSmithKline to kick-start a bid to become a specialist science college, a move that attracts increased government funding for equipment and staff. The schools spread these benefits by running evening classes, or introducing younger children to science. Taylor's moment with the little girl happened during an outreach visit to a local primary school.

Aisha Rama, a researcher in cardiac medicine, and plant scientist Steven Cook were among the first INSPIRE recruits. Both say that if they become teachers they are more likely to work with older pupils who have chosen to study science. Both saw INSPIRE as a route into teaching that allowed them to continue research projects while earning almost three times as much as trainee teachers.

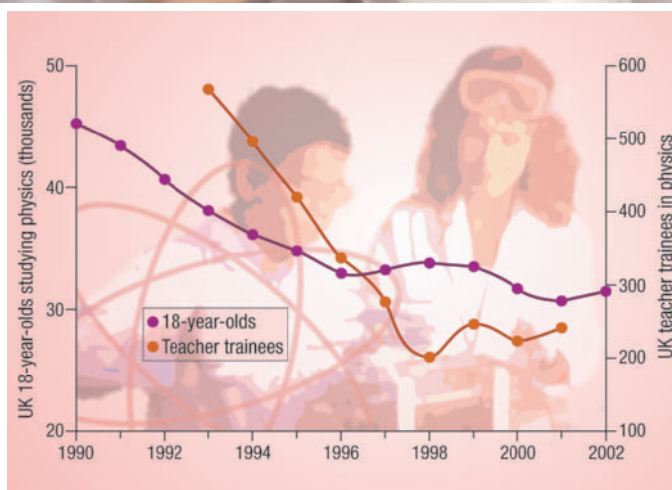
But juggling two jobs is a challenge. Even though their time should be split evenly between teaching and research, all the postdocs say that studying for a teaching qualification eats into their research time. It is virtually impossible to go to a conference if it happens during a teaching block, says Taylor.

Cook says his research adviser is totally supportive of the scheme, not least because he gets a free 'half-postdoc' for three years. Even though Cook's research has progressed slowly, he believes that having completed



INSTITUTE OF PHYSICS

Class act: researchers can make school science lessons more meaningful — and might help reverse the decline in numbers of physics students and teachers (right).



his teacher training he will be able to devote more time to the lab. It is certainly vital to have a supportive lab group: “The team is very adaptable, so the research can continue even when I’m not here,” says Litten.

Teacher training

Some postdocs are attracted to INSPIRE because it offers stability. “It’s a guaranteed bit of funding for three years, whereas a lot of postdoc jobs might only be for a year,” says David Chestnut, a laser physicist in his first year with INSPIRE. He is undecided about his future, but says that he enjoys teaching and is picking up transferable skills. “It’s certainly noticeable in public speaking. I went to a conference a couple of weeks ago where I was much more confident,” he says.

Taylor is candid that she saw this as a way to boost her chances of getting a university lectureship. “The one thing I didn’t have was a teaching qualification,” she says. She points

out that the UK government is pushing for all new lecturers to have a formal teaching qualification by 2006. “And if you want to learn teaching skills, it’s better to go to where the proper teaching is happening, which is in school,” she says. INSPIRE allows her to get a teaching qualification without dropping her research, and gives her an alternative career option that is easier to fit a family around. “It’s still a problem for female researchers to take a career break to have kids,” she says.

With only six postdoc-teachers so far, can INSPIRE make a real difference? Taylor thinks so: “If you just inspire one Nobel prizewinner then it’s been worth it — where would we be if Einstein had opted for English instead of science?” she says. “At first, our interest in INSPIRE was the money,” admits Ralph James, head of science at St Gregory’s High School in London, where Taylor teaches, but he was surprised by how much difference the postdocs made to practical

lessons. “If it’s a success then it could be a model to be adopted elsewhere,” he adds.

The University of Leeds is launching an equivalent scheme in September. Ceri Nursaw, head of the university’s city and regional office, is coordinating the scheme. She hopes that it will yield more applications for Leeds’s mathematics and science courses. And, she adds, going into schools is good for researchers: “Teaching will give them a better understanding of their subject.” It took Leeds two years to get going, but Nursaw thinks other universities could get a programme started in less than a year.

Shirley Ann Jackson, president of the American Association for the Advancement of Science, has campaigned for similar outreach efforts in the United States for more than two years. She argues that university or industry researchers could make a significant teaching contribution during the school year and focus on their research in the summer. But she says there are as yet no US schemes to foster scientist-teachers.

School’s out?

The key is money: everyone likes these initiatives, but nobody wants to pay for them. After receiving government funding during its first year, INSPIRE’s future is looking shaky. Imperial is now funding the post-docs’ research themselves, which Sykes says is unsustainable. Despite support from Britain’s science minister, David Sainsbury, Sykes says that “the good intentions just get lost in the bureaucratic machine. Kafka had nothing on this.” The dual nature of a post-doc’s work means that the government departments responsible for research and education each pass the buck, he adds.

Sykes also claims that the research councils that usually fund UK scientists are unwilling to contribute because of the unusual format of the research contracts.

Nor will GlaxoSmithKline guarantee its involvement beyond the trial period. “Who knows where we’ll be in two years time?” says Katie Pinnock, the company’s director of UK corporate contributions. But Imperial plans to hire more INSPIRE postdocs over the next two years, says Sykes, and support them until the end of their contracts.

There won’t be a shortage of applicants. In its first year, the scheme attracted more than 50 applications from Imperial for just three positions. All six postdocs say that they would recommend the scheme to anyone interested in teaching, and the schools remain keen. “The students have really warmed to the fact that these people are actually using science for something useful,” says Gareth Cross, a teacher at Stewards School in Harlow, Essex, and one of the science teachers involved. “And on top of that, they can bring a little bit of awe and wonder into the classroom.” ■

Mark Peplow is a reporter for Nature’s online news team.

It's life... isn't it?

Scientists find it hard enough to pin down evidence of early life on our own planet. How on Earth do we plan to determine whether life exists elsewhere? John Whitfield finds out.



We can, it's fair to say, be confident that there's life on Earth — but proving it is a different matter. In December 1990, the Galileo spacecraft pointed its sensors back at Earth before setting off for Jupiter. The probe reported an atmosphere abundant in oxygen and unusually rich in methane. It also detected a mysterious pigment that was unlikely to be of mineral origin: something earthlings call chlorophyll. Yet the astrophysicist Carl Sagan and his colleagues were still cautious in their conclusions based on these results. "Together, these are strongly suggestive of life on Earth," they wrote¹.

Galileo, admittedly, was not designed to detect life. And it did have a rather distant view of our planet, with only hints from the atmosphere to work out what was going on at the surface. But some dedicated, close-up examinations have proven no more conclusive. Efforts to detect faint traces of life in Earth's oldest rocks, some 4 billion years old, and in martian rocks that have fallen to Earth, have likewise produced results that are both ambiguous and disputed; claims of errors in procedure and interpretation fly back and forth with regularity.

Nevertheless, space agencies are shifting

their research agendas towards looking for life, says chemist Richard Mathies of the University of California, Berkeley. "Exobiology is becoming more prominent," he says. NASA and the European Space Agency (ESA) are moving from rock and mineral experiments to biological ones, he explains. In 2009, both agencies will send landers to Mars that, unlike the Spirit and Opportunity rovers now on the planet, will be designed to look for the chemistry of life.

Astrobiologists are currently working out what to look for, designing the instruments that could detect it, and using the most inhospitable, Mars-like environments on Earth as training grounds. Even so, these missions are highly unlikely to give a definitive answer. For that, we will have to wait at least another decade, when planned missions will bring back martian rocks for study in terrestrial laboratories.

The first experiments to look for life on Mars were the two Viking landers of 1976. Each lander scooped up a spoonful of soil from the planet's surface, moistened it with pure, sterile water, incubated it, and watched for evidence of chemical processes similar to those found in terrestrial microbes — such as the uptake of nutrients or exchange of gases.

As a control, the landers did the same tests on a heat-sterilized sample of martian soil.

Remarkably, every test gave a positive result. The soil samples released oxygen and compounds made from ingredients in the nutrient solution, and carbon in the experiment seemed to be incorporated into organic molecules. Unfortunately, most of the controls — except the experiments designed to detect nutrient uptake — also gave positive results. At the same time, two instruments designed to analyse the planet's chemistry, a gas chromatograph and a mass spectrometer, failed to detect any organic compounds on the surface of Mars².

Hit and miss

In sum, these experiments gave us some clues about the chemical environment on Mars, but failed to confirm or refute the existence of life. "The experiments all turned out to have some element of ambiguity, but that's utterly understandable," says chemist John Kerridge of the University of California, San Diego. Most researchers conclude that inorganic chemical reactions, perhaps powered by ultraviolet light, must have produced the chemicals detected³. A few still argue that the positive

NASA/SPL



Spot the difference: the Atacama Desert (above) is very similar to the martian landscape (left), which makes it an ideal place to test equipment for probing Mars for signs of life.

moon Europa, or Saturn's moon Titan.

Astrobiologists still face the problem highlighted by Viking: how do you ensure that 'evidence' of life isn't something produced by non-living processes? Such concerns have been at the heart of the debate on ALH84001, a 4.5-billion-year-old martian meteorite found in Antarctica in 1984. A team led by David McKay of NASA's Johnson Space Center in Houston, Texas, argued in 1996 that several features of the rock might have been produced by living organisms⁴. These were organic compounds, minerals similar to those produced by Earth bacteria, and structures resembling fossil bacteria.

But in the years following the publication, these claims came under sustained criticism. Some features, such as the proposed microfossils and mineral structures, could have been produced by non-living chemistry^{5,6}. Others, such as the organic chemicals present in the rock, may have been added to it after it arrived on Earth⁷. Few researchers now believe that ALH84001 is good evidence for the presence of past life on Mars.

Yet even those who disagree with the initial claims made for ALH84001 are glad of the effect the study had on the field. "Exobiology came out of the closet with that paper," says Mathies. "Suddenly it was OK to be working on this stuff." Christopher McKay and his colleagues also helped to define the categories of evidence — organic chemicals, minerals and fossils — necessary to build a convincing case for martian life, says Steele. "Even if the answers were wrong, the approach was good — they raised the bar by several metres."

Similar, and sometimes acrimonious, debates swirl around the evidence for the oldest life on Earth. Whether or not an outcrop of rock in Greenland contains evidence of life from 3.8 billion years ago is hotly contested⁸. And the question of whether structures in

3.5-billion-year-old rocks in Australia are fossilized bacteria or artefacts produced in hot springs has been controversial for several years^{9,10}. Those astrobiologists not actively involved in early-Earth research keep an eye on its developments, hoping that it can provide a blueprint for their field. "The issue of whether these microfossils have a biological origin or not is so contentious; it really shows the need to attack the problem from as many angles as possible," says Jorge Vago, study scientist on ESA's 2009 ExoMars project.

Both morphological and chemical signals are needed to confirm a claim. But both have their problems. "We've not yet identified a smoking gun for life," says geologist Jack Farmer of Arizona State University, Tempe.

A rocky road

The most recent attempt to look for evidence of life on Mars, the ill-fated Beagle 2 lander that presumably crashed on the planet last Christmas, was equipped to dig into the martian soil, to escape the surface environment that destroys organic compounds, and study the isotopic ratio of different elements. On Earth, living things preferentially incorporate the lighter form of carbon. The ratios of different isotopes of sulphur, iron, nickel and chromium have also been proposed as signs of biological activity.

But, again, it is not clear how geology and chemistry influence isotope ratios, particularly in rocks that may be billions of years old. Some believe that the carbon in Greenland's ancient rocks has a volcanic, rather than a biological origin^{11,12}. Geological processes that transform carbonate minerals into organic matter seem to produce compounds that, like biological remains, are enriched with the lighter form of carbon, says geochemist Mark van Zuilen of the Petrographic and Geochemical Research Centre in Nancy, France. "I don't think that carbon isotope ratios are that definitive an indicator of life," he says.

"We probably don't know all the ways in which isotope fractions can come about," adds geologist John Parnell of the University of Aberdeen, UK. "There may be things going on that we haven't thought of yet." Parnell is leading a UK project to identify biological molecules on the early Earth and Mars. His favoured biomarker is a class of organic molecules called hopanes, which are produced from the breakdown of simple cell walls. They are very long-lived — on Earth they persist for billions of years — and we know of no inorganic process that can produce them.

Mathies and his colleagues, on the other hand, believe that a good starting point is to look for amino acids. "If you look at life on Earth, 50% of the mass of biological materials is amino acids," he points out. These molecules can be made by inorganic processes as well as organic ones, and they have been

results should be taken as evidence for living microbes.

Ambiguity wasn't the study's only problem. Even if Viking had delivered a definite 'no' to life, this could simply have been because its instruments were too crude. "There could have been 10 million bacteria per gram of martian soil, and Viking wouldn't have seen them," says Andrew Steele of the NASA Astrobiology Institute at the Carnegie Institution of Washington.

Facing such problems, interest in looking for life on Mars waned after Viking, says astrobiologist Christopher McKay of the NASA Ames Research Center in Moffett Field, California. "The experiments were very hard to follow up on," he says. But there has been a boom in instrumentation since then — today's detectors can spot a single cell, or molecules at concentrations of parts per trillion. And our understanding of the tenacity of life in extreme environments has grown. On Earth, evidence for bacterial life has been found everywhere from deep oceans devoid of light, to hot springs at temperatures above boiling point. By the mid-1990s, biologists were once more on a mission to determine whether life could exist elsewhere in the Solar System — particularly on Mars, Jupiter's

"We've not yet identified a smoking gun for life."

— Jack Farmer



found on comets and meteorites, which might confuse matters. But it is important not to be too ambitious at first, says Mathies: after going to the expense of getting to Mars, you want to give yourself a reasonable chance of finding something. "You want to start general and get more specific," he says.

And there may be more telling evidence available from amino acids — on Earth, all biological amino acids have a certain geometry, whereas those not made by organisms are evenly split between two mirror-image conformations. In space, a bias either way could be a good indicator of biology at work. Amino-acid detectors are also more sensitive than isotope measurements, says Mathies. His team has proposed an instrument for NASA's 2009 Mars Science Laboratory (MSL) project that could pick this up.

Other biomolecules that might be worth looking for in space include nucleic acids and sugars. Most researchers favour a search strategy based on the chemistry of terrestrial life, but going slightly broader. For example, life elsewhere might use different nucleotides or amino acids from life here, or favour a different geometry or bias of isotopes. But structure and patterns in any of these areas would all be pointers to life. "Complexity is the most interesting biomarker," says Steele.

Of course, no mission would limit itself to pursuing one particular line of investigation. Future Mars landers will brandish a range of different tests, aided by lab-on-a-chip technology that has made instruments much smaller, as well as more sensitive. The ExoMars lander, for example, will carry cameras and microscopes, subject rocks to spectroscopic and chromatographic analysis to gauge their chemical composition and, if they look promising, grind them up to look for amino acids and other organic chemicals.

Animal, vegetable or mineral? Possible signs of ancient life in rocks in Greenland (above) and on the martian meteorite ALH84001 (inset) remain controversial.

It will also carry an antibody microarray chip that can detect extant life — or contamination by microbes carried from Earth¹³.

NASA's effort will be slightly different. "The MSL is very much a chemical lab: it won't directly look for life, but maybe for the residues or precursors of life," says Firouz Naderi, head of the Mars programme at the agency's Jet Propulsion Laboratory in Pasadena, California. The MSL is a scouting party for the proposed 2013 Mars sample-return mission and a 2016 astrobiology field laboratory; it may even hoard rocks for these landers to collect.

Out on the edge

To test all these sensitive instruments, researchers are currently digging through the dirt of Earth's harshest ecosystems. Parnell and his colleagues have been looking for organic chemicals in the Haughton impact crater, the relic of a past meteor impact in the Canadian Arctic, where NASA tests much of its Mars science and technology. Steele's team also spent last summer in the Arctic — this time in Svalbard, where the rocks share some geological features with ALH84001 — testing their life-spotting equipment on the local microbes. His team hopes to get a life-detection-chip on the ExoMars lander.

Christopher McKay and Mathies, mean-

while, are both working in the Atacama Desert in Chile. "It's the only place on Earth where there's no life at the surface at all," says McKay. He spent June looking for the residues of organic compounds on desert rocks, and digging holes to see if bacteria live underground there. The researchers work in special suits to avoid swamping any microbes in the soil with their own bacteria.

Another collaboration, between NASA and the Center for Astrobiology in Madrid, Spain, is looking for life, and testing the technology needed to drill and sample on Mars, in the acidic, mineral-rich Rio Tinto river in southern Spain. The acid is made by bacteria, which metabolize sulphur and produce sulphuric acid. Carol Stoker of NASA Ames, the project's leader, was delighted when the Opportunity rover landed on a plain of sulphate minerals on Mars that likewise seem to be the remains of an acidic sea. "It was a slam dunk," she says. "The chemistry of the

martian surface is exactly like what Rio Tinto is producing now."

The researchers plan to test an automated drilling platform next spring, and also have an instrument, called SOLID (signs-of-life detector), that they hope to get on one of the 2009 missions.

So, despite the occasional false start and blind alley, the astrobiologist's tool kit and knowledge base is becoming ever broader. Most researchers are now confident that, if there is or has been life elsewhere in the Solar System, we have the ability to detect it. But to do so, we will still have to be lucky enough to study the right rock in the right place with the right techniques — and get it safely back to Earth. "For a biosignature to be believed, it has to be returned for a sample," says Kerridge. "Reproducibility convinces hard-nosed scientists." Even then, don't bet on everyone being convinced. ■

John Whitfield is a science writer based in London.

1. Sagan, C., Thompson, W. R., Carlson, R., Gurnett, D. & Hord, C. *Nature* **365**, 715–721 (1993).
2. Biemann, K. & Lavoie, J. M. *J. Geophys. Res.* **B 84**, 8385–8390 (1979).
3. Plumb, R. C., Tantanayon, R., Libby, M. & Xu, W. W. *Nature* **338**, 633–635 (1989).
4. McKay, D. S. *et al.* *Science* **273**, 924–930 (1996).
5. Bradley, J. P., Harvey, R. P. & McSweeney, H. Y. Jr *Nature* **390**, 454 (1997).
6. Barber, D. J. & Scott, E. R. D. *Proc. Natl Acad. Sci. USA* **99**, 6556–6561 (2002).
7. Jull, A. J. T., Courtney, C., Jeffrey, D. A. & Beck, J. W. *Science* **279**, 366–369 (1998).
8. Dalton, R. *Nature* **429**, 688 (2004).
9. Brasier, M. D. *et al.* *Nature* **416**, 76–81 (2002).
10. Schopf, J. W., Kudryavtsev, A. B., Agresti, D. G., Wdowiak, T. J. & Czaja, A. D. *Nature* **416**, 73–76 (2002).
11. Fedo, C. M. & Whitehouse, M. J. *Science* **296**, 1448–1452 (2002).
12. Van Zuilen, M. A., Lepland, A. & Arrhenius, G. *Nature* **418**, 627–630 (2002).
13. Clarke, T. *Nature* **413**, 247–248 (2001).

Ocean noise could injure more than mammals

Geologists should wait until more is known about the harm their work may do to fish.

Sir — Your News story “Push to protect whales leaves seafloor research high and dry” (*Nature* **428**, 681; 2004) reports that the Lamont-Doherty Earth Observatory survey of the 65-million-year-old Chicxulub meteorite crater, coordinated by the US National Science Foundation (NSF), was cancelled because of concerns that the airguns used could harm marine mammals.

As a science adviser for Seaflow — a conservationist organization with a mission to reduce harmful underwater noise (see www.seaflow.org) — I am concerned that the only instrument we have to protect the ocean environment resides in legislation protecting marine mammals.

Lamont-Doherty did apply for marine mammal ‘incidental harassment

authorization’, and received authorization from the National Marine Fisheries Service to deploy the airguns under the rules of the US Marine Mammal Protection Act. The Chicxulub survey was ultimately blocked by a Mexican environmental agency that was worried about marine mammals, but I believe that there are other good reasons to refrain from seismic airgun research until we know what effects the blasts have on the marine environment.

What is not included in US law, or even in the discussion, is the impact that airgun explosions have on fish populations.

There is evidence that seismic airguns can damage fish ears (R. D. McCauley *et al.* *J. Acoust. Soc. Am.* **113**, 638–642; 2003). We also know that fisheries are compromised by airgun operations. The Chicxulub

survey was to take place in 3,000 km² of tropical waters at depths of 50–100 m. To ignore the probable impact that this study would have had on fish populations reveals an added dimension to the short-sightedness of the programme.

Although halting the survey may have been costly in the short term, it may make more sense to spend NSF money looking at the impacts of anthropogenic noise on fish. Once the biological research is more complete, geologists can do their research using methods that will not compromise the biota, raise the ire of fishermen and conservationists, or force action by environmental protection agencies.

Michael Stocker

Seaflow, 1062 Fort Cronchite, Sausalito, California 94965, USA

Edwardian anaesthetists had a finger on the pulse

Sir — I read with great interest in your 100 Years Ago column (*Nature* **428**, 705; 2004) an excerpt from the 14 April 1904 issue of *Nature* which addressed the increasing death rates under chloroform anaesthesia. Back then, the American and British schools of anaesthesia had chosen to use different general anaesthetics. The Americans had embraced diethyl ether, with its more forgiving margin of error, allowing administration by physicians, nurses or medical students. The British preferred chloroform, despite its narrower margin of error, which led to more frequent complications (including death) and required a physician to administer it safely. The ongoing controversy in America over whether nurses or physicians should administer anaesthetics is, arguably, left over from this early difference in practice.

The irony is that the 1904 author criticized “the ignorant and careless anaesthetist” for using ‘cardiac syncope’ (a loss of consciousness or fainting due to the heart not pumping enough blood to the brain) as an excuse following death under chloroform. However, the most likely cause of death in patients who died suddenly under chloroform anaesthesia was — we now know — a ventricular fibrillation or tachycardia, causing the patient to immediately become pulseless and unconscious. So cardiac syncope was probably the correct postmortem diagnosis, even though the physiological basis for the syncope could not have been fully appreciated until decades later.

So how did the anaesthetists of 1904 make such a reasonable diagnosis? Before the introduction of a cuff for measuring blood pressure, and recording it on a written graph, the anaesthetist’s left hand customarily held the mask to the patient’s face, while his left little finger continuously registered the pulse under the patient’s chin. This could be skilfully done, after some training, while maintaining a firm mask-grip. The anaesthetist would therefore know immediately when the pulse faded. Indeed, this manual skill is still useful, whenever our microprocessor-controlled blood-pressure devices fail during a mask anaesthetic, and we must know the patient’s status. We also derive the expression, “to keep one’s finger on the pulse”, from this anaesthetic technique long predating the Edwardian era.

Francis X. Dillon

Department of Anesthesia, Indiana University School of Medicine, Fesler Hall Room 204, 1120 South Drive, Indianapolis, Indiana 46202-5115, USA

Supplementary materials need the right format

Sir — It is becoming increasingly common for articles in top-tier journals to be accompanied by online supplementary text, figures, graphs or other materials. These supplementary materials typically provide either additional evidence for an article’s claim or a detailed description of the methods used to carry out the study. Unfortunately, the format in which this material is stored is not always optimal.

Nature’s policy is essentially “anything

goes”, accepting supplementary files in a wide array of formats, and publishing them as provided by the author (see www.nature.com/nature/submit/finalsubmission/SI/index.html). *Nature’s* preferred format for text, tables and images is the Microsoft Word format (.doc), which seems to be the natural choice, given the widespread use of Microsoft Word by scientists.

Unfortunately, .doc is particularly ill-suited for archival and online-publishing purposes. Whether a particular .doc file can be opened and printed successfully depends on the exact version of Microsoft Word installed, the version of the operating system installed, the printer installed, and the fonts installed. Furthermore, the details of the .doc format are secret and change from version to version. As a result, some of *Nature’s* readers will have problems opening and printing supplementary material. Moreover, we should expect that many of these documents will fail to open properly 10 to 20 years from now.

I believe that by allowing publication of .doc files alone, *Nature* does a disservice both to its authors and to its present and future readers. At the minimum, *Nature* should convert .doc files into portable document format (.pdf) — which has fully published specifications (see <http://partners.adobe.com/asn/tech/pdf/specifications.jsp>) and is suitable for online publishing and archiving — and publish the .pdf files alongside or instead of the .doc files. Similar procedures have been adopted by *Science* and by the *Proceedings of the National Academy of Sciences*.

Claus O. Wilke

Keck Graduate Institute of Applied Life Sciences, 535 Watson Drive, Claremont, California 91711, USA

The man who knew power

A look at one of the key figures in the development of the atomic bomb.

Edward Teller: the Real Dr Strangelove

by Peter Goodchild

Weidenfeld & Nicolson: 2004. 352 pp. £25

To be published in the US by Harvard University Press

Barton J. Bernstein

More than twenty years ago, Peter Goodchild, a television producer at the BBC, wrote a book about J. Robert Oppenheimer, father of the atomic bomb. Now he turns his attention to the man who became Oppenheimer's foe, the nuclear-weapons physicist and policy adviser Edward Teller, who died last year. The book on Teller is a useful, although error-prone, popular history that should help to guide the writing of more scholarly, probing, careful and better volumes. In the interim, this engaging biography will stand as the best single book on Teller's whole life.

The book is largely fluent, often gripping, sometimes insightful and occasionally poignant — in short, it is a good read. However, it sometimes understates its dependence on other published works, notably Gregg Herken's *Brotherhood of the Bomb* (Henry Holt, 2002), which devotes about 100 pages to Teller.

One of the most controversial US scientists of the post-war years, Teller was a Hungarian emigré of Jewish descent who went to the United States in 1935. He is best known to the public and much of the science world not for his formal contributions to theoretical physics, but for his efforts on, and disputes about, nuclear weapons and nuclear policy. These included his campaign for the hydrogen bomb, his negative testimony in 1954 when Oppenheimer was accused of disloyalty, his opposition to the test-ban treaty, and his crusades for civilian nuclear power and for the US Strategic Defence Initiative (SDI), or 'star wars' programme.

Despite the book's possibly polemical subtitle, Goodchild usually weaves a path between the views of Teller's admirers and his enemies, though tilting more to the side of unfriendly critics. He does not directly challenge Teller's sincerity, but occasionally laments his pursuit of power and the compromises, in Goodchild's view, that such a quest presumably propelled Teller to make.

Many studies of Teller note the troubling discrepancies between the archival (and sometimes public) record and Teller's own claims about major issues and events. Among others, these include Teller's different attitudes before and after Hiroshima on the use of the atomic bomb, his many



Don't look back: Edward Teller denied having been in favour of dropping the bomb on Hiroshima.

dealings with Oppenheimer, his testimony in the Oppenheimer loyalty hearing, and his excessively optimistic reports about the SDI. But unlike most analysts, Goodchild avoids making firm judgements on whether these discrepancies have their roots in lies or self-deception, or perhaps some other kind of innocent error.

While discussing these discrepancies, Goodchild quotes his recent interview with Edward Teller's son, Paul, a philosophy professor at the University of California, Davis. Paul Teller says his father was the most honest man he knows, and concludes that the emotional quality of some disputes distorted his father's recollections, producing honestly held, but possibly inaccurate, memories. Edward Teller's foes would obviously be less charitable, and some quoted interviews in the book point in that direction.

The book deals uneasily with the troubling issue of Teller's pre-Hiroshima attitude about using the atomic bomb on Japan in 1945. He actually supported combat use of the bomb, apparently without a prior non-combat demonstration to warn Japan of its power. Unfortunately, because of inadequate research, Goodchild does not recognize how greatly Teller struggled in the aftermath of Hiroshima and Nagasaki to rewrite his own pre-Hiroshima past on the bomb's use.

To help understand Teller generally, Goodchild uses pieces of Teller's frequently fascinating correspondence in the 1940s and 1950s with physicist Maria Goeppert Mayer, who won a Nobel Prize in 1963 for discoveries concerning nuclear shell structure. Quoting from those sometimes obliquely flirtatious letters, Goodchild skill-

fully captures arresting glimpses of the self-doubting, argumentative, spirited, brooding, wary and embattled Teller. With Maria, he often played the role of the wayward imp and made her the adult scold.

Goodchild, despite missing some important themes and evidence, stresses how much Teller was injured by his testimony against Oppenheimer in the 1954 disloyalty case when fellow scientists learned of his harsh words. Goodchild poignantly describes the deep depression into which Teller fell. According to Goodchild, Nobel laureate Enrico Fermi feared that Teller might be close to suicide in 1954 after the backlash to his testimony.

Teller's support for the SDI occupies a few chapters, which often lean heavily on the writings of journalist William Broad. Some of Goodchild's interviews provide previously little-known details about the SDI but otherwise leave the story largely unchanged.

Teller survived to the age of 95, outliving nearly all his scientific contemporaries. Becoming heavily dependent upon others in illness and old age, he hired students to read to him. Some of them, in a matter apparently undiscovered by Goodchild, had no inkling of who Edward Teller was. To them he was simply an old man, nearly blind and in poor health, who paid them to read aloud, even though he sometimes fell asleep during the session.

Goodchild's book, like Herken's, ends rather evocatively and provocatively. In 1987, Teller, dissenting from Oppenheimer's 1940s statement that the physicists after the dropping of the atomic bomb had known sin, concluded that they have instead

BETTMANN/CORBIS

“known power”. Was Teller driven, and corrupted, by the quest for power? Goodchild suggests that he was.

Despite such a tantalizing but generally underdeveloped theme, the book only rarely probes beneath the surface of Teller’s political and personal life. It frequently relies too heavily on interviews to illuminate much earlier events and ignores many archival materials and some relevant secondary literature. There are also numerous errors (I spotted more than 50) in discussing events and quoting materials, describing people’s careers and spelling names, and citing titles and authors.

Vigorously scrubbed and with full sourcing, including explicit reliance on other scholarship, this readable biography would deserve the significant audience it may well gain among those who want an inviting survey of Teller’s life. It tells a story of Teller’s personal and policy battles, with notable defeats and memorable victories. ■

Barton J. Bernstein is in the Department of History, Stanford University, Stanford, California 94305-2024, USA.

Gentle biases

Biased Embryos and Evolution

by Wallace Arthur

Cambridge University Press: 2004. 248 pp.
£50, \$85 (hbk); £18.95, \$32 (pbk)

Armand M. Leroi

This book is an introduction to the principles of ‘evo-devo’ — evolutionary developmental biology. It is written with exemplary clarity and charm, and is clearly intended for the general reader or undergraduate. Beginning with a balanced account of the various strands of modern evolutionary thought, it goes on to outline the fossil history of animals, fruitfly developmental genetics, phenotypic plasticity, phylogenies and the various ways in which genes and ontogenies can change over evolutionary time.

So far, so conventional, even boring. But make no mistake. The author is Wallace Arthur and, as with all his books, there is a whiff of sulphur about this one. It comes on page 13 when he discusses orthogenesis, the notion that lineages have an intrinsic drive to evolve in particular directions. “Orthogeneticists were seen by many as mystics,” Arthur writes. “But, even though I have no time for mysticism, I have to admit some sympathy for their cause.”

These are brave words. Orthogenesis has been a cause without mainstream sympathizers for at least 60 years. The reason for this is that no one has provided a mechanism by which it might work. Most biologists believe that the evolutionary direction of lineages is largely determined by natural selection;



A question of size: can biases in mutation rate alter the direction of evolution in different animals?

a minority make great play of contingency (the non-selective effects of meteor strikes and the like). So what is going on? Has Arthur discovered a new principle of evolution?

Not really, no. The fuel in his orthogenetic engine is ‘mutation bias’. Mutation produces novel phenotypes, but it does not produce all novel phenotypes in equal frequency in a given population. For example, mutations that cause an animal to become smaller than normal might be more common than those that cause it to become larger. This bias is the result of the way body size is specified in development — a bias that might influence the direction that evolution takes, causing small animals to evolve more often than large ones.

To epitomize Arthur’s position, there is “a bias in the production of variant phenotypes or a limitation on phenotypic variability caused by the structure, character, composition or dynamics of the developmental system”. The quote isn’t from his book, it is from John Maynard Smith’s famous position paper (*Q. Rev. Biol.* **60**, 265–287; 1985) defining what most of us call ‘developmental constraints’ — it’s just that Arthur doesn’t like the term. Many, perhaps most, evolutionary biologists accept that developmental constraints exist. If they aren’t a major topic of study — and they should be — it is because distributions of mutational effects are very hard to measure.

Mutation bias is not enough to produce orthogenesis, however. If there is a single fitness optimum, or if the population is sufficiently large to ensure that all possible mutations are always present, then the direction of evolution will be dictated by natural selection alone. But if the landscape is rugged and population sizes small, the particular peak climbed by a population could depend

on what mutations happen to be available. This is not orthogenesis of old — which posited a force independent of, or even capable of opposing, natural selection — but a reassignment of influence over evolutionary trajectories from natural selection to the kind of genetic variation available for it to work on.

If ‘mutation bias’ turns out to be a new term for an old idea, the same seems to be true for another unusual term: ‘internal selection’. This is the idea that as one part of an organism evolves, it exerts selective pressure on other parts to change as well. Suppose a mutation increasing the length of an animal becomes fixed in a population. This might cause the subsequent fixation of another mutation that increases the animal’s width, so restoring an original, harmonious, proportion. Arthur makes great play of this, but I think the interaction at the heart of this process is well known to population geneticists as ‘fitness epistasis’ and has often been experimentally demonstrated.

Evolutionary biologists will not be convinced by Arthur’s arguments, for they are quite free of both data and maths. But some formal theory (for example, L. Y. Yampolsky and A. Stoltzfus *Evol. Dev.* **3**, 73–83; 2001) does underpin his claims, and it should force us to consider the relative influence of mutation and natural selection on evolution more carefully than we might have done. But this book is ultimately meant for general readers. They will find it a gentle and engaging account of how modern developmental genetics is beginning to affect the neodarwinian agenda. ■

Armand M. Leroi is in the Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

The story of structure

Chasing the Molecule

by John Buckingham

Sutton: 2004. 259 pp. £20

Alan Rocke

One of the least familiar of all the heroic stories in the history of science is how chemists managed to learn about the invisible realm of atoms and molecules. At the dawn of the nineteenth century, few scientists paid attention to the world at the molecular level. It was beyond the reach of the senses and it seemed to many at the time that there was little we could ever know about it, aside from the sort of metaphysical conjectures that had engaged philosophers since antiquity.

By the end of the century, the situation had changed radically. Chemists were now in a position to depict the detailed architecture of complex molecules, often even being able to specify the manner in which the atoms were arrayed in three dimensions — all without ever having been able to see these sub-microscopic objects. By 1900, certain branches of physics operated on the nano-scale, too, but it was the chemists who had blazed the trail. The fascinating story of how they did it is filled not only with interest, but also with drama and romance.

John Buckingham, a respected organic chemist not previously known for historical writing, recounts this engrossing history for a general audience in *Chasing the Molecule*. After providing three chapters of historical background, he introduces John Dalton, the first true chemical atomist. Buckingham then proceeds through the famous names and events of nineteenth-century chemistry: the imperious Swede Jöns Jacob Berzelius; the gentle sage of Göttingen, Friedrich Wöhler; the irascible German Justus Liebig; the domineering Parisian Jean-Baptiste-André Dumas; the romantic radicals Auguste Laurent and Charles Gerhardt; the obscure Scot Archibald Couper; the late defenders of organic radicals, Edward Frankland and Hermann Kolbe; the great master of chemistry in London and Berlin, August Wilhelm von Hofmann; and the father of structure theory, August Kekulé.

Buckingham's exposition is clear and lively, his ability to explain the scientific action on stage is first rate, and his command of historical detail is impressive. The task he set for himself is not one for the timid. The story he tells is complex, filled not only with difficult science but with many false starts, misunderstandings, confusing complications, dead ends and long chains of inference that often had weak links. Even among professional historians of science, few are closely

Exhibition

Fond memories

The discovery of the photographic process in the nineteenth century revolutionized the sciences as well as the creative arts. Natural-history illustrators were among the first to experiment with 'drawing with light', yet this brief movement, centred on a group of mostly English and Scottish artists, has faded from conscious memory.

Anna Atkins' 1853 *Cyanotypes of British and Foreign Flowering Plants and Ferns* was the first printed book to include photographs and text. Atkins created her cyanotypes, such as the alga *Laminaria digitata* illustrated here, by laying specimens directly on to light-sensitive paper.

Atkins' work on seaweed species features prominently in the exhibition *Ocean Flowers: Impressions from Nature in the Victorian Era*, which can be seen at the Yale Center for British Art in New Haven, Connecticut, until 8 August. This is a rare chance to see a visual history of Victorian botanical illustration.

A book to accompany the exhibition,



Ocean Flowers, edited by Carol Armstrong and Catherine de Zegher, is published by Princeton University Press (\$49.95, £32.95). **A.A.**

familiar with this terrain; the specialist literature is large but of uneven quality. To his credit, Buckingham has read (and cites) a sizeable amount of this secondary literature, including many of the most reliable sources. There are occasional minor errors (consistently misspelling the names Chevreul and Sertuerner, for instance), but considering all the complexities, it is truly impressive how much detail Buckingham has got right.

The author's eye for interesting stories and arresting metaphors gives the book real appeal and bite. For example, in explaining the historical emergence of valency, as distinct from chemical 'affinity', Buckingham notes the difference: "An atom can have powerful urges but be satisfied with a single partner, like Queen Victoria, or polygamous but sluggish, like her uncle George IV."

Occasionally, however, Buckingham rests his interpretations on shaky foundations, for instance when coming to his surprisingly mixed judgements on the careers of what are in my opinion the three greatest chemical masters of the nineteenth century (or any other): Berzelius, Liebig and Kekulé. During his prime in the 1820s and 1830s, Berzelius had a well deserved reputation as the "lawgiver of chemistry"; in his old age (he died in 1848) he exerted a stultifying influence on the development of chemical

theory, but it is unjust to focus on that time in his career to the exclusion of his extraordinary earlier accomplishments. Liebig is a paradox. Often generous and amiable, at other times he rode roughshod over the reputations of other chemists. His overall contributions to the science were undeniably massive, however. And some who have attacked Kekulé's stories of inspiration through dreams (which I believe are fully credible) have gone so far as unjustly to impugn his character. It is tempting for any historian to gravitate towards the titillating controversies, but one needs to be careful to maintain balance and perspective.

Professional historians may disapprove of the author's propensity to assess blame, to deplore his protagonists' "errors" and to judge past scientific contributions by today's standards, but they are not the principal audience for the book. In fact, in recounting the story of the rise of chemical atomic theory for a general audience, the author has accomplished what few specialists have even attempted. I hope that this book will be read widely; the author deserves congratulations for his skilled recounting of a complicated and important story. ■

Alan J. Rocke is professor of history at Case Western Reserve University, Cleveland, Ohio 44106, USA.

Blood traffic control

Red blood cell vasodilation: nitric oxide and haemoglobin help to match blood flow to metabolic demand.

David J. Singel and
Jonathan S. Stamler

Textbook ideas regarding the delivery of oxygen to the body's tissues tend to focus on the cooperativity in the binding and release of oxygen from the transport protein haemoglobin. The uptake of oxygen in the lungs and release of oxygen to the tissues are coupled to structural changes in the haemoglobin tetramer that synergistically alter its affinity for oxygen ligands. Although this mechanism is important, its significance is overemphasized when looking at how oxygen reaches tissues — the main determinant of oxygen delivery is blood flow.

In the microcirculation, blood flow is regulated by physiological oxygen gradients. The progressive decrease in the oxygen content of blood (the oxygen saturation of haemoglobin) as the diameter of the arteriolar blood vessels decreases is coupled to graded vasodilation. A decrease in the oxygen content of blood, in response to (for example) increased muscle activity, stimulates an increase in blood flow. When the demand for oxygen is reduced, the oxygen content of the blood rises and blood flow decreases. Thus, blood flow is precisely matched to metabolic requirements (tissue oxygen consumption). The detailed biochemical mechanism by which oxygen evokes these vascular responses remains unknown.

Arterial blood flow — the trafficking in oxygenated red blood cells — is governed by local modulation of the diameter of small resistance vessels, suggesting a role for the vasodilator nitric oxide (NO). It should be noted that the NO regulation of vessel diameter that resides within the endothelium, and is dependent on shear stress and chemical agonists, is separate from the NO regulation of blood flow that occurs within red blood cells, and is dependent on oxygen. But the sensor of blood oxygen content and the mechanism by which the oxygen signal is transduced to modulate NO bioactivity are unknown. Haemoglobin seems to be an ideal oxygen sensor, but prevailing ideas about the chemistry of its interaction with NO present a conceptual roadblock.

Haemoglobin, in both its oxygenated and deoxygenated forms, was known to react with NO long before there was any awareness of NO's vasoactivity. But as neither of the resulting NO-derived species directly induce vasodilation, it was thought that haemoglobin would only quench NO bioactivity. The discovery of a previously unknown



In the flow: red blood cells in the microcirculation.

compound produced with NO, an S-nitrosothiol derivative of haemoglobin that has potent vasodilatory activity, changed the picture. Most significantly, this activity is potentiated by the deoxygenation of haemoglobin. Haemoglobin can therefore sense ambient oxygen levels through its oxygen-binding function and adjust NO bioavailability. In turn, red blood cells can cause microvessels to dilate or constrict, with S-nitroso-haemoglobin as an intermediate. Thus, the outlines of a mechanistic basis for matching tissue perfusion with metabolic demand were at hand.

Although numerous studies verified the formation of S-nitroso-haemoglobin after NO/haemoglobin interaction, others found that NO was bound and rendered inactive by haemoglobin. These divergent results can be understood by considering both the emergent complexity of haemoglobin — well appreciated in the realm of oxygen binding, but rarely considered in its overall NO biochemistry — and the undervalued importance of physiological conditions. Although the use of excess NO relative to haem was justified in early experiments, where NO was used primarily as a surrogate for oxygen, it gives a false picture of the physiological situation, in which the ratio is 1 in 1,000–10,000 and oxygen content is a critical variable.

At physiological concentrations, NO reacts preferentially with the β -subunits of haemoglobin (each tetramer contains two α - and two β -subunits), forming S-nitroso-haemoglobin in the oxygenated state. In contrast, at supra-physiological levels, NO tends to lodge on the α -subunits and favour the deoxygenated state. The disposition and reactivity of NO bound to haemoglobin is thus a function of multiple variables, including the other allosteric effectors of haemoglobin (pH; partial pressure of carbon dioxide, $p\text{CO}_2$; partial pressure of O_2 , $p\text{O}_2$) and the ratio of

their concentrations to haem. Changes in conditions can give rise to stark, rather than subtle, linear changes in the distribution of NO reaction products. This complexity emerges both from the cooperative, nonlinear behaviour of the haemoglobin tetramer and from the branched network of coupled kinetic equations underlying this rich chemistry.

How are these ideas relevant to oxygen delivery under physiological conditions? In the blood perfusing tissues, oxygen is in great excess over NO. Haemoglobin is also in excess over other blood oxides of nitrogen with which it interacts (for example, nitrite). Under such conditions, NO binds preferentially to the β -subunit, with S-nitrosothiol formation promoted at higher $p\text{O}_2$ and NO group release promoted at lower $p\text{O}_2$. Thus, S-nitroso-haemoglobin contributes to the matching of blood flow to demand under physiological conditions. In contrast, when micromolar concentrations of NO arise, as in septic shock, the potential problem of excess S-nitroso-haemoglobin and consequent excessive vasodilation is avoided by sequestering NO on the α -haems, which additionally lowers the overall oxygen affinity of the protein. This chemistry restricts NO bioavailability while enhancing oxygen delivery.

A complete kinetic/thermodynamic model of NO's interactions with haemoglobin in the circulation — a respiratory cycle for NO — has yet to be achieved. But an appreciation of the complexity of this problem is essential in designing experiments to elucidate the detailed mechanism by which red blood cells act as oxygen-responsive transducers of NO bioactivity. Emerging evidence that vasodilation by red blood cells is altered in disease, including heart failure, pulmonary hypertension and diabetes, should open a new field of investigation and could potentially change the practice of medicine. ■

David J. Singel is in the Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana 59717, USA. Jonathan S. Stamler is at the Howard Hughes Medical Institute, Duke University Medical Center, Durham, North Carolina 27710, USA.

FURTHER READING

- Duling, B. R. & Berne, R. M. *Circ. Res.* **37**, 325–322 (1970).
- Jia, L. *et al. Nature* **380**, 221–226 (1996).
- González-Alonso, J. *et al. J. Physiol.* **530**, 331–341 (2001).
- Datta, B. *et al. Circulation* **109**, 1339–1342 (2004).
- Singel D. J. & Stamler J. S. *Annu. Rev. Physiol.* **67**, (in the press).

A three-player solution

Lewi Stone

The seemingly unpredictable 'boom and bust' of insect-pest populations will be better understood with the advent of a deceptively simple model combining field and laboratory data with earlier theories.

Etienne Leopold Trouvelot, an astronomy professor at Harvard in the nineteenth century, had his scientific career "ruined by a moth"¹. As it turns out, Trouvelot, an obsessive amateur entomologist, could not have chosen a more formidable opponent. In 1868, his experiments with the gypsy moth (*Lymantria dispar*) resulted in disaster after several of the insects escaped from his suburban Boston home. The moth was an alien species. It proceeded to multiply, only slowly at first, but some 20 years later it could be found in its millions, defoliating forests and causing major economic and ecological damage as it went on to invade North America (Fig. 1).

Enormously embarrassed, Trouvelot, the classic 'good man gone wrong', returned to France and to obscurity.

Alas, the same can't be said of the gypsy moth, which remains one of the most devastating forest pests to this day. Pinning down the ecological processes that give rise to outbreaks of this type of insect pest, and modelling their complex dynamics, has troubled theoretical ecologists for decades: in this issue Dwyer *et al.*² (page 341) present a mathematical model that might just crack the problem.

Broadly speaking there are two competing theories to explain insect outbreaks. The first argues that climatic factors, such as rainfall and temperature, set appropriate conditions for outbreaks. The other suggests that biotic interactions — competition and predation, for instance — are largely responsible. Mathematical models based on experimental data can provide insights about the extent to which these different factors are at play. Beyond this, however, there are some strange circumstances in need of an explanation. For example, of the approximately 80 different forest species of moths and butterflies (Lepidoptera) that exhibit large outbreaks, at least 18 are to some degree cyclical. That is, outbreaks recur in a regular fashion, and although not perfectly cyclical they have an average periodicity of 8–14 years³.

It is still not understood why these patterns have an underlying cyclicity, although one would expect there to be some basic



Figure 1 Bare birch — an upshot of the gypsy moth invasion of eastern North America. These grey birch trees would be in leafy splendour had they not been stripped of developing foliage. Inset, a gypsy moth caterpillar, the insect stage that does the damage.

force that drives it. The recurrent patterns, moreover, seem to be highly resilient to perturbation and do not break down with external manipulation: the cycling seems to possess some form of ecological stability. On top of all this, the insect outbreaks exhibit remarkable spatial synchrony, with populations rising and falling simultaneously, sometimes across millions of square kilometres.

Early theoretical explanations for insect outbreaks were based on the idea of multiple equilibria. According to the theory, density-dependent processes maintain the population at low levels for long periods. In the case of the gypsy moth, small-mammal predators such as the white-footed mouse regulate the population at low density until such time as the predators themselves drop in number because of the random failure of other food sources. The gypsy moth population then bursts into outbreak mode and rapidly jumps to the higher equilibrium level. At these higher densities, pathogens, especially a nuclear polyhedrosis virus (NPV), rapidly infect the population, causing enough gypsy moth mortality to trigger the collapse of the outbreak⁴. This, however, is only part of a complex story. A more comprehensive

description should also include the so-called induced-defence hypothesis, whereby the deteriorating quality of forest foliage due to defoliation has a negative impact on the gypsy moth population.

In the 1970s, insect-outbreak models were designed to mimic some of these features by incorporating multiple equilibria with alternative stable states at low and high densities. These models, however, did not include the critical ingredient of host–pathogen dynamics. In contrast, host–pathogen-type models of the 1980s and 1990s generally failed to take account of the density-dependent regulating effects of predation. Dwyer *et al.*² are the first to glue these two processes together, by careful and direct inclusion of the interactions between all three players — the gypsy moth, its predators and its specialist pathogen, NPV. In doing so, their model is able to reproduce long-term time series of gypsy moth outbreaks with greater success than any other so far. Moreover, the basic framework should prove equally applicable to modelling the dynamics of many other pests that defoliate forests.

The model of Dwyer *et al.* seems deceptively simple, but there is more to it than meets the eye. First, it is based on more than a decade of meticulous field and laboratory investigations of gypsy moth host–pathogen

NATURAL RESOURCES CANADA, CANADIAN FOREST SERVICE

transmission dynamics⁵. These studies not only provided well defined estimates of ecological parameters, but also identified an essential ingredient in outbreak dynamics, namely that gypsy moth larvae vary greatly in their susceptibility to the virus. Heterogeneity of this form has a known stabilizing effect and is a crucial contributing factor in ensuring the recurrence of outbreaks. Dwyer and colleagues' model takes this heterogeneity into account, along with the essential time delays that occur between larval infection and actual death, as well as inter-epidemic dynamics such as pathogen survival from one season to the next, and differences in mean susceptibility among larval stages. The version of the model analysed here is, in fact, the 'infinite epidemic/burnout approximation' of a more sophisticated set of equations⁵, but it elegantly captures all the main features.

Dwyer *et al.* note that their host-pathogen model inherits the multiple equilibria of the 1970s models. But, in addition, it generates even more complex, and sometimes chaotic, dynamics. With the inclusion of environmental stochasticity, to mimic the effects of weather, the model population jumps unpredictably between multiple attractors (different nonequilibrium 'attracting' states), producing time series of changes in population density that closely resemble those of real gypsy moth populations. Model outbreaks occur on average every 11 years, but with a variability from outbreak to outbreak that closely matches real data. Between outbreaks, the predator maintains the model gypsy moth population at relatively low levels. With time and stochasticity, however, the population slowly builds up to trigger the next outbreak. On inclusion of spatial structure, the authors show that different subpopulations synchronize their chaotic outbreak oscillations and remain relatively in phase with one another, which is in keeping with the coupled oscillator theory from nonlinear dynamics⁶. This result helps to explain the enigmatic spatial synchronization of many pest species.

Now that Dwyer *et al.*² have exposed the critical mechanisms behind gypsy moth outbreaks, it would be interesting to take their work further. For example, the model would allow assessment of the relatively new and effective natural enemy of the gypsy moth, the fungal pathogen *Entomophaga maimaiga*, present in North America since 1989. Unlike NPV, which triggers the collapse of a gypsy moth outbreak, the fungus can sometimes prevent outbreaks altogether. In some areas of North America, *E. maimaiga* has the potential to bring gypsy moth damage to a halt. Modelling work (see, for example, ref. 7) is currently under way to assess the dispersal dynamics of the fungus, its ability to become established and its overall impact. Modifications of Dwyer and

colleagues' model might help in this regard. Their model will undoubtedly prove of great value for studying the dynamics of many other pest species that cause episodes of ecological devastation. ■

Lewi Stone is in the Biomathematics Unit, Department of Zoology, Tel Aviv University, Ramat Aviv 69978, Israel.
e-mail: lew521@yahoo.com

Imaging techniques

Seeing single spins

P. Chris Hammel

Combining the imaging power of magnetic resonance and the sensitivity of atomic force microscopy has created a hybrid technique that can resolve single spins beneath the surface of a sample.

Magnetic resonance imaging (MRI) and atomic force microscopy (AFM) are two of the most powerful imaging technologies available. MRI can provide, non-invasively, fully three-dimensional images from deep within an object. However, the spatial resolution of this technique is limited to the smallest volume that contains enough nuclear or electronic spins to generate a detectable signal. Sensitivity to a smaller number of spins is the key to improving spatial resolution.

Combining three-dimensional MRI with the excellent force sensitivity of AFM — in magnetic resonance force microscopy, or MRFM — opens the possibility of performing scanned-probe MRI with much improved spatial resolution. On page 329 of this issue, Rugar *et al.*¹ report the combination of MRI and AFM to achieve sensitivity to a single electron spin. Compare this with the 10^8 – 10^{10} spins required in a conventional electron-spin resonance experiment. This signal achievement will dramatically alter the horizons for high-resolution imaging.

But the ability to detect individual spins is about more than imaging — it implies the power to manipulate individual spins as well. Present-day information processing relies on the electron's charge, through manipulating and detecting voltages in electronic circuits. Exploiting the electron's magnetic moment, or spin, could lead to significant enhancements in electronic information processing, including nonvolatile memory, increased integration densities and reduced power consumption². Furthermore, the spin of the electron is a natural two-state quantum system ('qubit') for quantum computing; the spin can also be isolated from its physical environment to achieve the long decoherence times needed for successful computation.

MRI exploits the proportionality between the easily measured frequency of a magnetic resonance signal and the value of the magnetic field at the spin's location. In

1. Liebholt, A., Mastro, V. & Schaefer, P. W. *Bull. Entomol. Soc. Am.* **35**, 20–21 (1989).
2. Dwyer, G., Dushoff, J. & Yee, S. H. *Nature* **430**, 341–345 (2004).
3. Myers, J. H. *Adv. Ecol. Res.* **18**, 179–242 (1988).
4. Elkinton, J. S. & Liebhold, A. M. *Annu. Rev. Entomol.* **35**, 571–596 (1990).
5. Dwyer, G., Dushoff, J., Elkinton, J. S. & Levin, S. A. *Am. Nat.* **156**, 105–120 (2000).
6. Blasius, B., Huppert, A. & Stone, L. *Nature* **399**, 354–359 (1999).
7. Weseloh, R. M. *Biol. Control* **29**, 138–144 (2004).

an applied magnetic field, the resonant frequency of spins increases steadily across a region, following the increasing field. Thus measuring the resonant frequency pinpoints the location of the spins responsible for the signal. And because magnetic fields penetrate samples easily, three-dimensional images can be constructed from signals from deep inside them. MRI has had a huge impact in the biomedical arena, but there is continuing demand for higher spatial resolution, below the 1 mm^3 currently achievable in medical settings.

A decade ago, John Sidles proposed³ a radical approach to MRI, based on detecting the force exerted by the spins in a sample on a microscopic magnet, which is mounted on a flexible cantilever above the sample. This would offer the much improved sensitivity needed to reduce the imaged volume and to achieve atomic-scale nuclear MRI⁴. Rugar and colleagues' demonstration¹ of single electron-spin detection using Sidles' approach is a heartening milestone in realizing the dream of high-resolution MRI. Two elements are essential for the dramatically improved sensitivity: large magnetic-field gradients generated by micrometre-scale magnetic probe tips, and highly sensitive cantilevers. Rugar *et al.* engineered cantilevers with a state-of-the-art force sensitivity⁵ of 10^{-18} newtons specifically for this purpose.

The micromagnetic probe mounted on the cantilever tip generates a large field gradient (approximately proportional to its magnetization divided by its diameter). In this set-up¹ (Fig. 1), the gradient of the microscopic magnetic probe was approximately 2 gauss per nanometre, so that the force generated on the cantilever by an individual electron-spin was detectable, at 2×10^{-18} newtons. The field gradient has a second, independent role: as in MRI, it causes spins located at different depths beneath the micromagnetic tip to resonate at different

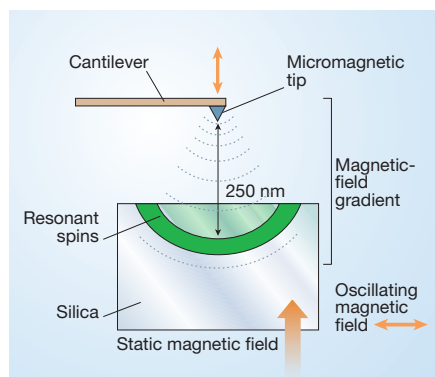


Figure 1 Magnetic resonance force microscopy. A micromagnetic tip is mounted on a specially engineered cantilever above a silica sample containing a low density of electron spins. In the presence of the large field gradient of the tip, the applied oscillating magnetic field excites electron spins at a particular depth in the sample at their resonant frequency. The force exerted by a single electron through its magnetic moment on the cantilever, although only at the level of 10^{-18} newtons, can be detected — as shown by Rugar *et al.*¹, who have used this technique to achieve the detection of a single electron spin.

frequencies; this provides the basis for selective excitation of spins, and hence imaging. Among scanned-probe techniques, MRFM is uniquely able to image definable volumes beneath a surface. In Rugar and colleagues' experiment, 250 nm separates the tip and the resonant spin buried in the sample below. The large field gradient and the low spin density in the sample (of irradiated vitreous silica) ensure that individual spins can be selected.

This accomplishment also built on crucial advances in understanding and control of electron spin relaxation. The visibility of weak signals can be enhanced by coherently adding them together, because noise contributions, being random, will tend to add to zero. This improvement is limited by the lifetime of signal coherence, so long spin lifetimes are advantageous. However, bringing the micromagnetic tip close to the spins can reduce lifetimes by speeding relaxation⁶: thermal fluctuations of the cantilever move the micromagnetic tip randomly, so spins near the tip experience fluctuating magnetic fields; these fluctuations cause spins to relax⁷. Rugar *et al.* controlled this problem by fabricating a mass-loaded cantilever in which the problematic cantilever tip motions are suppressed⁸.

MRFM is a unique single-spin detection technique, in that it couples directly to the electron's magnetic dipole moment. Other approaches exploit mechanisms that couple the spin of the electron to its spatial degrees of freedom, allowing detection through coupling to its charge (for instance, by monitoring an optical transition that is sensitive to the spin state^{9,10}). Another approach rests on

the Pauli exclusion principle, which sets symmetry constraints on the electron wave function (the mathematical description of its quantum state). As the electron wave function is a product of its spatial and spin components, the Pauli principle effectively couples spin and spatial degrees of freedom. Charge is much easier to detect than spin, but the better sensitivity afforded by these techniques — through charge detection, instead of spin detection — comes at the price that only in particular situations is it possible to arrange the required interaction between electron spin and orbital degrees of freedom.

In contrast, the MRFM approach is entirely general and can, in principle, be applied to the detection of any magnetic moment (nuclear¹¹ and ferromagnetic¹² MRFM have also been demonstrated). The 'in principle' qualification is a consequence of the difficulty of detecting an individual spin; this limits how rapidly measurements can be made. Further improving sensitivity could greatly enhance the range of applicability of MRFM. Higher spatial resolution will benefit fields from nanoelectronics to biomolecular imaging. As electronic elements become smaller, they become sensi-

tive to the presence of individual impurities and dopants; hence three-dimensional atomic-scale characterization becomes crucial. Perhaps most notably, the field of spin-based quantum computation will require single-spin detection technology both for the readout of quantum states and for device characterization; single-spin MRFM promises to aid in this challenging undertaking.

P. Chris Hammel is in the Department of Physics, Ohio State University, Columbus, Ohio 43210, USA.

e-mail: hammel@mps.ohio-state.edu

1. Rugar, D., Budakian, R., Mamin, H. J. & Chui, B. W. *Nature* **430**, 329–332 (2004).
2. Wolf, S. A. *et al. Science* **294**, 1488–1495 (2001).
3. Sidles, J. A. *Appl. Phys. Lett.* **58**, 2854–2856 (1991).
4. Sidles, J. A. *et al. Rev. Mod. Phys.* **67**, 249–265 (1995).
5. Stowe, T. D. *et al. Appl. Phys. Lett.* **71**, 288–290 (1997).
6. Stipe, B. C., Mamin, H. J., Stowe, T. D., Kenny, T. W. & Rugar, D. *Phys. Rev. Lett.* **87**, 096801 (2001).
7. Mozyrsky, D., Martin, I., Pelekhov, D. & Hammel, P. C. *Appl. Phys. Lett.* **82**, 1278–1280 (2003).
8. Mamin, H. J., Budakian, R., Chui, B. W. & Rugar, D. *Phys. Rev. Lett.* **91**, 207604 (2003).
9. Wrachtrup, J., von Borczyskowski, C., Bernard, J., Orritt, M. & Brown, R. *Nature* **363**, 244–245 (1993).
10. Kohler, J. *et al. Nature* **363**, 242–244 (1993).
11. Rugar, D. *et al. Science* **264**, 1560–1563 (1994).
12. Zhang, Z., Hammel, P. C. & Wigen, P. E. *Appl. Phys. Lett.* **68**, 2005–2007 (1996).

Embryology

Plane talk

Gerald Schatten and Peter Donovan

In mammals, is the three-dimensional body plan ingrained in the egg at or before fertilization? The answer is 'maybe, but then again maybe not'. Less invasive techniques might help to resolve matters.

Hans Spemann¹ neatly summed up the importance of embryonic axes for correct animal development: "We are standing and walking with parts of our body which could have been used for thinking had they developed in another part of the embryo." But how and when are the embryonic axes established? Elsewhere in this issue², Hiiragi and Solter (page 360) deliver the latest contribution to a debate over dividing embryos that is dividing embryologists.

Some unfertilized vertebrate and invertebrate eggs display undeniable polarity along a plane known as the animal–vegetal axis. The animal pole is defined by an area of meiotic cell division (whereby a nucleus divides into four, each resulting nucleus containing half of the original complement of chromosomes); by default, the opposite pole is vegetal. This axis is variously apparent from differences in surface smoothness and pigmentation, and features of the egg cytoplasm. But until recently such features had not been identified in mammalian eggs, raising the question of whether we mammals abandoned the instructions found in non-

mammalian eggs for assembly of our body's three-dimensional plan.

Just a few years ago, groundbreaking work by Richard Gardner and collaborators^{3,4}, as well as Magdalena Zernika-Goetz and her group^{5–7}, provided compelling evidence that mouse eggs possess polarities that track closely with the later body planes. Perhaps in mammals, then, as in other animals, one or more of the three body planes (anterior–posterior, dorsal–ventral and left–right) are already ingrained in the egg before or during fertilization. From their studies with time-lapse video microscopy, however, Hiiragi and Solter² assert that the mouse egg does not display any predetermined axes.

During the maturation of an egg cell, it undergoes by meiosis a unique asymmetrical division into two cells: the secondary oocyte (which will become the egg) and the polar body. Further meiosis results in the asymmetrical division of the secondary oocyte into the mature egg and the second polar body, each containing a single set of chromosomes. Normal development results in the eventual degradation of the polar bodies, but while they persist they are often used as land-

marks by those studying the structure of eggs.

Once the nuclei (pronuclei) of egg and sperm have combined, cell division in a fertilized egg or zygote is by mitosis — the division of single nuclei into two new nuclei, each containing a complete copy of the original cell's chromosomes. A fertilized egg divides mitotically into many smaller cells, in a cleavage process that is marked early on by the formation of a cleavage furrow at the surface of the egg. This furrow is usually precisely placed above where the genetic material is separating, ensuring that each daughter cell receives identical sets of chromosomes.

Gardner³ was the first to seriously ponder the use of the mouse egg's polar body as a navigational beacon. He and others⁸ typically found the polar body to be located in the first cleavage furrow. With Hiiragi and Solter's report, the reliability of polar-body positioning as a determinant of the cleavage plane is questioned. They report that although many mouse eggs divided within 30° of the second polar body, many did not. So they doubt whether the second polar body is a fiduciary mark of the cleavage plane, and whether unfertilized mouse eggs even have an animal-vegetal axis.

Since first witnessing the frog egg's 'grey crescent', which forms on the freshly fertilized egg's pole opposite to the sperm-entry point, embryologists recognized that gastrulation — the stage at which cell division and migration start to form a complex structure with a primordial central gut tube — later starts at this region. Piotrowska and Zernicka-Goetz⁴ reported that in mice the meridian created by the sperm entry point specified the second embryonic axis. Hiiragi and Solter, again using time-lapse 'ovo-vision' to trace the relations between the fertilization cone (the structure around the sperm entry point) and first cleavage, report that this meridian is a less accurate predictor of cleavage than the orthogonal orientation of the apposed sperm and egg pronuclei before they fuse into a single nucleus following fertilization.

For their ultimate experiment, Hiiragi and Solter scrambled mouse eggs by cloning early-stage male or female pronuclei into fertilized eggs from which they had previously removed the male pronucleus. Remember, the original question is whether an unfertilized egg has pre-existing asymmetry or whether the sperm entry site defines a critical meridian. So the rationale here is that transferring an early-stage male pronucleus might carry with it a signal for cleavage determination. Again, the orientation of the apposed male and female pronuclei was the most accurate predictor of the first cleavage plane, because this axis defines the axis of the first mitotic apparatus, with cytokinesis — the division of cytoplasm that follows nuclear division during mitosis — occurring

orthogonal to this structure.

Early cell biologists appreciated the importance of the pronuclei's journey. The starting position of the egg's meiotic spindle, the site of female-pronucleus formation and the sperm entry point all influence the trajectories of pronuclear migrations and the ultimate axis of pronuclear apposition. Cleavage occurs orthogonal to the mitotic spindle's axis and pronuclear positioning, just before breakdown of the nuclear envelope deposits the centrosomes, crucial intercellular architecture that helps to organize chromosomes, at their ultimate positions when the mitotic spindle assembles. Consequently, the trajectories of the sperm's entrance and the cytoplasmic migration of the two pronuclei, culminating in their apposition near the egg centre, will influence their final positions just before mitosis.

So perhaps the polar body is not always the precise marker of the egg's 'North Pole', in the way that magnetic north is off-axis from geographic north. And perhaps the sperm entry point, especially if restricted from the area over the meiotic spindle, provides only transient indications of directionality and not long-term reliability in that respect.

It could of course be that the strategies for observing this subtle, easily perturbed signal influence the experimental results and thereby undermine the conclusions drawn from them. A decade's worth of evidence comes from skilled teams, each repeatedly having performed exacting experiments. But, at least to us, each experimental set seems influenced by the second polar body's own

predilection to move into the furrow, as well as the strength of its tether to the egg surface — all potentially influenced by the egg's compression during imaging. Elegant studies injecting oil or fluorescence markers into the egg, or examining the movement of cytoplasm, are questioned by non-believers as artefacts, as are results from studies involving pronuclear extractions and subsequent nuclear fusions.

The results of Hiiragi and Solter² will force renewed attention on the natural forces that generate the embryo's three-dimensional complexities, and perhaps less invasive examinations of the processes involved. Meantime, we remain fascinated by the polarized views on polarity and polar bodies. ■

Gerald Schatten is in the Department of Obstetrics, Gynecology and Reproductive Sciences, University of Pittsburgh School of Medicine, 204 Craft Avenue, Pittsburgh, Pennsylvania 15213, USA.

e-mail: gschatten@pdc.magee.edu

Peter Donovan is in the Institute for Cell Engineering, and the Department of Gynecology and Obstetrics, Johns Hopkins University, Broadway Research Building, Suite 559, Baltimore, Maryland 21205, USA.

e-mail: pdonova4@jhmi.edu

1. Spemann, H. *Forschung und Leben* (Engelhorn, Stuttgart, 1943).
2. Hiiragi, T. & Solter, D. *Nature* **430**, 360–364 (2004).
3. Gardner, R. L. *Development* **124**, 289–301 (1997).
4. Gardner, R. L. *Development* **128**, 839–847 (2001).
5. Piotrowska, K. & Zernicka-Goetz, M. *Nature* **409**, 517–521 (2001).
6. Piotrowska, K. & Zernicka-Goetz, M. *Development* **129**, 5803–5858 (2002).
7. Plusa, B. et al. *Nature Cell Biol.* **4**, 811–815 (2002).
8. Weber, R. J., Pedersen, R. A., Wianny, F., Evans, M. J. & Zernicka-Goetz, M. *Development* **126**, 5591–5598 (1999).

Plant biology

Good neighbours

Laura Serna

Plants depend on structures called stomata to regulate gas exchange with the air, and their positioning is crucial. A key factor controlling stomatal development and arrangement has now been found.

We all need good neighbours, but plant guard cells need them more than most. Destined to remain where they are produced, these cells cannot function efficiently without help from adjacent cells. Guard cells are arranged in pairs to create stomata, and occur mainly on the surfaces of leaves; they surround a pore through which the plant exchanges gases with the atmosphere. The pore aperture is regulated by changes in the rigidity of the guard cell, and this in turn is controlled by the flow of water and ions between the guard cells and their neighbouring non-guard cells. Now, writing in *Science*, Bergmann and co-workers¹ report that a recently identified enzyme in the flowering plant *Arabidopsis thaliana*², christened

YODA (YDA), is crucial to the formation and arrangement of stomata.

Mutations in the YDA gene result in a plant in which almost all the cells at the plant surface are guard cells. As might be expected, this overproduction of stomata has severe consequences, and many of the mutant seedlings die, while those that survive produce small plants with sterile flowers. Mutations in three other genes — the catchily named TOO MANY MOUTHS (TMM)³, STOMATAL DENSITY AND DISTRIBUTION1 (SDD1)⁴ and FOUR LIPS (FLP)⁵ — also affect stomata, producing stomatal clusters, but also allowing the formation of single stomata. The single stomata ensure correct gas exchange and so these three mutants develop normally. But how

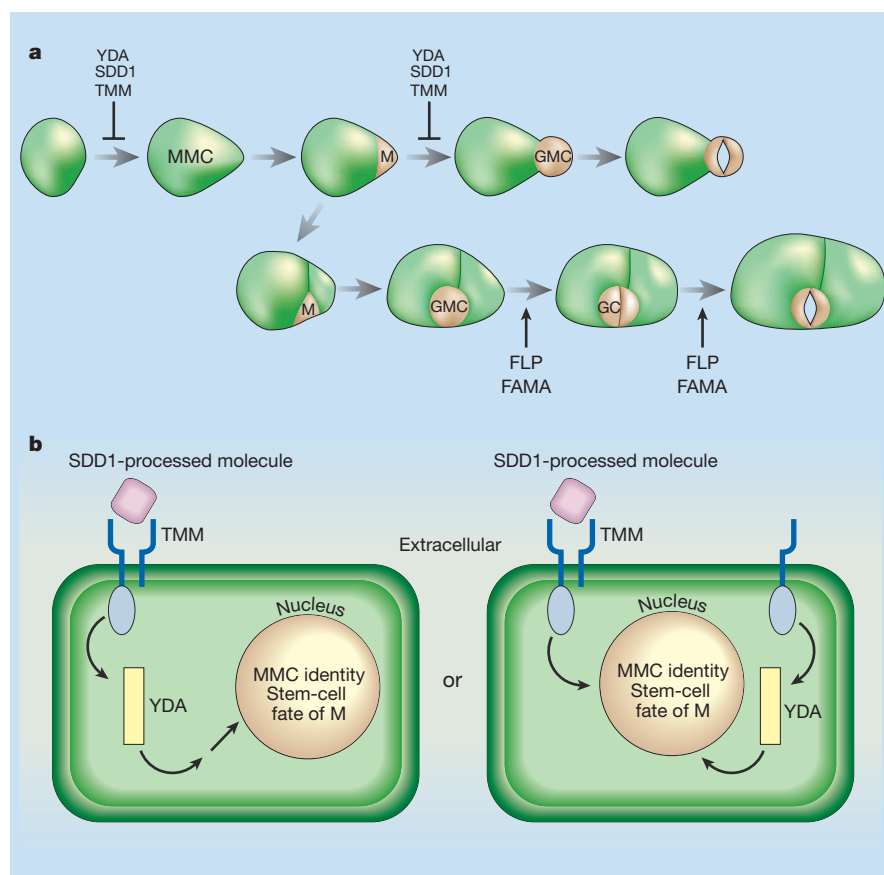


Figure 1 YODA and the development of stomata. **a**, Stomatal development begins with a meristemoid mother cell (MMC), which divides to produce a meristemoid (M) and a pavement cell. Meristemoids are self-renewing cells, producing new meristemoids. After a number of cell divisions they develop into guard mother cells (GMCs), which then divide to produce two guard cells (GCs). Bergmann *et al.*¹ suggest that YODA (YDA) promotes the MMC identity and maintains the stem-cell activity of the meristemoids. TMM and SDD1 also regulate such functions. The authors propose that FAMA controls GMC identity and GC differentiation, functions shared with FLP. **b**, YDA acts downstream of SDD1/TMM or in a parallel pathway. Left, an SDD1-processed molecule might activate a TMM/co-receptor complex. The hypothetical co-receptor would provide TMM with a cytoplasmic domain, which it lacks. Activation of this receptor complex would activate YDA, triggering a transduction pathway that initiates the stomatal-cell lineage and maintains the meristemoids' stem-cell activity. Right, activation of YDA might depend on an as-yet-unidentified receptor, and an SDD1/TMM/co-receptor might trigger a signal-transduction pathway independently of YDA. Both pathways would control the two functions above.

does YDA prevent such overproduction of stomata?

In *Arabidopsis*, guard-cell formation is the final step in a series, or lineage, of cell divisions that leads to the formation of the leaf surface⁵ (Fig. 1a). This process starts with a precursor cell called a meristemoid mother cell (MMC), which divides to form a small meristemoid and a larger pavement cell. The meristemoid is a self-renewing cell, but after a few divisions it loses this stem-cell activity and becomes a guard mother cell (GMC). The GMC then divides to form a pair of guard cells. In *yda* mutants, in which YDA activity is lacking, increased numbers of cells seem to become MMCs, which prematurely form stomata because the meristemoids that are produced assume a GMC identity. This suggests that YDA prevents stomata overproduction by reducing the number of cells

entering the stomatal lineage and controlling the balance between meristemoid renewal and stomata formation. Further studies may confirm these functions and might also identify new ones.

The protein encoded by YDA appears to be a MAPKK kinase, an enzyme that adds phosphate groups to proteins, thus regulating their activity. The putative YDA protein is very similar to members of the STE11 family of such kinases². MAPKK kinases are upstream components of so-called kinase signalling cascades, which link cell-surface receptors to critical regulatory targets within cells⁶. When not required, the enzymatic activity of such kinases is suppressed by a regulatory region of the enzyme molecule — the amino-terminal domain^{6,7}. Thus, YDA should be continuously switched on in plants expressing constructs of the enzyme

in which the N-terminal regulatory domain has been deleted (ΔN -YDA plants). Strikingly, Bergmann and colleagues¹ find that ΔN -YDA plants totally suppress guard-cell formation. The authors infer from this that entry into the stomatal lineage requires a reduction in YDA activity. Deletion of this regulatory region also causes a loss of signalling specificity⁶, but the characteristics of the *yda* mutants indicate that the absence of stomata in ΔN -YDA plants is due to the continuous activation of the YDA pathway.

SDD1/TMM and YDA may share at least two functions — regulating entry into the stomatal lineage and maintenance of the meristemoids' stem-cell activity. Do they act in the same genetic pathway? Bergmann *et al.* show that a single copy of the ΔN -YDA gene, which has no effect in wild-type control plants, can restore normal stomatal production in both *tmm* and *sdd1* mutants. The authors argue that YDA acts either downstream of SDD1/TMM or in a parallel pathway. The structure of these molecules seems to suggest the former possibility, and that the receptor protein TMM might react with a molecule processed by SDD1^{5,8}, thereby activating YDA (Fig. 1b). Because TMM has no kinase domain linking it to the cell interior, it might need to dimerize with a co-receptor that would provide such a domain^{5,8}. In fact, the presence of paired cysteine amino acids in TMM suggests that it forms disulphide bridges with a partner^{5,8}. The resulting receptor complex would transduce the signal from the cell surface to the cell interior, where YDA is activated.

The ability to generate plants with contrasting features — *yda*, in which all the surface cells are guard cells, and ΔN -YDA, completely lacking guard cells — provides an opportunity to uncover genes that might regulate stomatal development. Bergmann *et al.* take matters a step further by comparing all the genes expressed in *yda* and ΔN -YDA seedlings. They find that the expression of 111 genes is reduced in ΔN -YDA and increased in *yda*. As might be expected, TMM and SDD1, which are expressed in stomatal precursor cells or stomata^{8,9}, fall into this group. Bergmann *et al.* also find that the expression of a further 109 genes is increased in ΔN -YDA and reduced in *yda* mutants. Downstream targets of YDA should be in this group. Because MAPKK kinases usually act in more than one signalling cascade, thereby regulating several plant features⁶, only some of these genes should control stomatal formation and arrangement.

But that's not all. A recent study of *Arabidopsis* identified mutants for 74% of the plant's genes¹⁰, which should greatly accelerate study of the function of the 220 genes identified by Bergmann *et al.* The authors have followed this course and find that mutations in a gene called FAMA produce clusters of early-stage stomata that are reminiscent of

the stomatal clusters seen in the *flp* mutant³. One exciting consideration, however, has not been emphasized here. FAMA belongs to a family of gene-transcription factors that have a 'basic helix-loop-helix' structure¹ and act with MYB proteins to regulate several processes in *Arabidopsis* development¹¹. Will the long-awaited *FLP* gene encode (or regulate) an MYB protein?

Bergmann and colleagues' study not only uncovers a new function for YDA, it also provides a powerful tool for identifying genes that control stomatal development and pattern formation. Further studies are needed to define the gene's function and to identify the components of the YDA cascade. Undoubt-

edly, we have a long but exciting way to go, but also the tools to take us there. ■

Laura Serna is in the Facultad de Ciencias del Medio Ambiente, Universidad de Castilla-La Mancha, E-45071 Toledo, Spain.
e-mail: laura.serna@uclm.es

1. Bergmann, D. C., Lukowitz, W. & Somerville, C. R. *Science* **304**, 1494–1497 (2004).
2. Lukowitz, W. *et al.* *Cell* **116**, 109–119 (2004).
3. Yang, M. & Sack, F. D. *Plant Cell* **7**, 2227–2239 (1995).
4. Berger, D. & Altmann, T. *Genes Dev.* **14**, 1119–1131 (2000).
5. Serna, L. & Fenoll, C. *Trends Genet.* **18**, 597–600 (2002).
6. Hirt, H. *Proc. Natl Acad. Sci. USA* **97**, 2405–2407 (2000).
7. Asai, T. *et al.* *Nature* **415**, 977–983 (2002).
8. Nadeau, J. A. & Sack, F. D. *Science* **296**, 1697–1700 (2002).
9. Groll, U. V. *et al.* *Plant Cell* **14**, 1527–1539 (2002).
10. Alonso, J. M. *et al.* *Science* **301**, 653–657 (2003).
11. Schiefelbein, J. *Curr. Opin. Plant Biol.* **6**, 74–78 (2003).

Solar physics

Hidden magnetism

Jan Olof Stenflo

Observations of the Hanle effect have revealed the existence of small-scale 'hidden' magnetic flux on the quiet Sun. The magnetic-energy density of this hidden flux is much larger than previously thought.

Magnetic fields have occupied centre stage in solar physics for the past several decades, and have come to be regarded as the key ingredient for a unified understanding of solar phenomena. It may therefore come as a surprise that, after all these years, the magnetic-energy density in the solar atmosphere might have been seriously underestimated — as Trujillo Bueno *et al.*¹ conclude on page 326 of this issue.

The spectrum of radiation from the Sun can be resolved into a series of lines corresponding to different atomic transitions. In the presence of a magnetic field, these spectral lines can split into multiple polarized components — this is the Zeeman effect and it has been used to diagnose the magnetic fields on the Sun since it was first introduced² to astrophysics by George Ellery Hale in 1908. Ever smaller magnetic structures have been revealed, and there is no end in sight. In fact, the magnetic flux seems to have a fractal structure, with an almost scale-invariant, self-similar pattern³ (Fig. 1). According to theoretical predictions based on magnetoturbulence, the structuring should continue for several orders of magnitude beyond the scales that have so far been resolved.

But the Zeeman effect as a diagnostic tool is 'blind' to magnetic fields that are tangled on scales too small to be resolved. Below the scale of the achievable angular resolution, the contributions to the Zeeman-effect polarization from opposite-polarity components within a tangled magnetic field cancel each other; so an unresolved mixed-polarity field leaves no 'footprint' in the Zeeman-split

spectral lines. For this reason, a vast amount of solar magnetic flux has possibly remained hidden from view.

Trujillo Bueno *et al.*¹ have used a different approach. The alternative tool is the Hanle effect, a coherence phenomenon discovered⁴ by Wilhelm Hanle in 1924. This effect was of great significance in the early development of quantum mechanics, because it demonstrated the principle of the coherent superposition of quantum states, and the nature of decoherence when the degeneracy of the quantum states is partially lifted as weak magnetic fields are introduced⁵. In the context of solar physics, the Hanle effect refers to the set of polarization effects that are caused by the coherent scattering of the radiation in the Sun's atmosphere under the influence of an external magnetic field. The symmetry properties of the Hanle effect with respect to the orientation of the magnetic field are entirely different from those of the Zeeman effect. Thus an unresolved, tangled magnetic field leaves a polarimetric footprint for the Hanle effect, while being invisible to the Zeeman effect.

Although the Hanle effect was introduced decades ago as a tool to gain information on the elusive, turbulent solar magnetic field⁶ (and a lower limit of 10 gauss on the turbulent field strength was determined straight away), further progress was limited by the insufficient polarimetric precision of the instruments available. The polarization amplitudes sought are small (typically at the level of 0.1% or less), because the anisotropy of the radiation field in the solar atmosphere is so small. But the wealth of polarization phenomena caused by the

coherent scattering of the Sun's radiation at last became fully accessible with the introduction a decade ago of the Zurich Imaging Polarimeter, ZIMPOL⁷, with which the polarimetric noise could be dramatically reduced. Using ZIMPOL, the electro-optically modulated polarization signal is 'demodulated' into four image planes, which correspond to the different polarization states needed to form the full Stokes vector (which contains the complete polarization information).

With the polarimetric accuracy of one part in 10⁵ that is routinely reached with ZIMPOL, combined with high spectral resolution, an astounding variety of spectral structures can be seen throughout the whole solar spectrum. This linearly polarized spectrum — which has been called the 'second solar spectrum', because it bears so little resemblance to the ordinary intensity spectrum — is a veritable treasure trove of all kinds of coherence phenomena⁸. The different polarized structures in the second solar spectrum are affected to various degrees, through the Hanle effect, by the 'hidden' magnetic fields in the solar atmosphere. Differential effects can then be used as diagnostics.

However, the proper quantitative interpretation of the Hanle signatures in the solar spectrum is a tricky business. The polarization amplitudes observed unfortunately depend on details of how the spectral line is formed — including the details of the structure of both the atom/molecule and of the solar atmosphere, as well as the way in which the polarized radiation is transported through the atmosphere. Trujillo Bueno *et al.*¹ have brought the theory of line formation to a new level: they have modelled the Hanle effect for spectral lines from atoms and molecules using a three-dimensional radiative-transfer code and a model of the Sun's atmosphere obtained from simulations of the star's surface convection. They find much higher magnetic-energy densities than in previous investigations, both because of their more realistic three-dimensional approach and because they introduce a more realistic probability distribution function for the turbulent field strengths, rather than using a single-value field.

This work is a pleasing example of how different areas of astrophysics can be brought together to achieve a scientific objective: Trujillo Bueno *et al.* have combined high-precision spectro-polarimetry, coherency effects in atomic physics, advanced techniques in radiative-transfer theory and simulations of magnetoconvection. The information that can be retrieved about the Sun's magnetism is, however, model-dependent, and will remain so for the foreseeable future, because the small-scale magnetic structures in question will not be resolved even with the next generation of

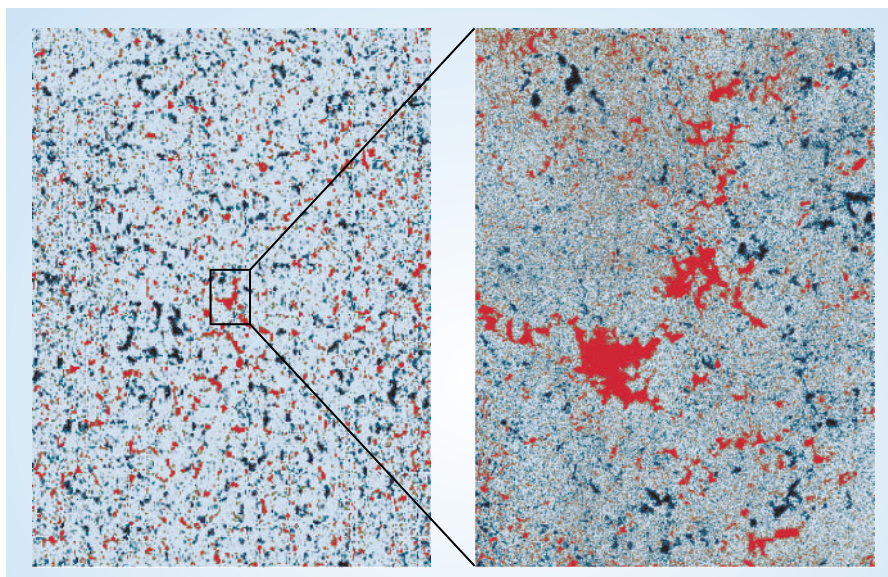


Figure 1 The fractal-like pattern of magnetic fields on the quiet Sun. The enlarged image on the right (courtesy of Göran Scharmer, from the Swedish Solar Telescope on La Palma) covers 1% of the area of the left panel (from the Kitt Peak Observatory). These maps represent the patterns of circular polarization in the Sun's radiation caused by the Zeeman effect. The red and blue patches represent flux of opposing magnetic polarities, separated by the grey voids of seemingly no flux. Observations made using the Hanle effect now reveal¹ that these grey regions are not voids at all, but are teeming with turbulent fields that carry a significant density of magnetic energy.

telescopes. Still, with the wealth of new diagnostic information available through the second solar spectrum, we can expect to see major advances in our understanding of the Sun's hidden magnetism in the years to come.

Jan Olof Stenflo is at the Institute of Astronomy, ETH Zürich, CH-8092 Zürich, Switzerland.
e-mail: stenflo@astro.phys.ethz.ch

1. Trujillo Bueno, J., Shchukina, N. & Asensio Ramos, A. *Nature* **430**, 326–329 (2004).
2. Hale, G. E. *Astrophys. J.* **28**, 315–343 (1908).
3. Janssen, K., Vögler, A. & Kneer, F. *Astron. Astrophys.* **409**, 1127–1134 (2003).
4. Hanle, W. Z. *Phys.* **30**, 93–105 (1924).
5. Moruzzi, G. & Strumia, F. (eds) *The Hanle Effect and Level-Crossing Spectroscopy* (Plenum, New York, 1991).
6. Stenflo, J. O. *Solar Phys.* **80**, 209–226 (1982).
7. Povel, H. P. *Opt. Eng.* **34**, 1870–1878 (1995).
8. Stenflo, J. O. *et al. Astron. Astrophys.* **355**, 789–803 (2000).

Developmental biology

Heading away from the rump

Ray Keller

How does our rump come to be separated from our head, instead of being right behind our ears? Studies of the elongation of the developing embryo reveal some remarkable underlying mechanisms.

The early vertebrate embryo consists of a relatively featureless ball of thousands of cells, which reshapes itself to form a head at one end, a rump or tail at the other, and an elongated trunk in between. Much of this reshaping is accomplished by a group of cells called chordamesodermal cells. In the early embryo, they form a wide but short area just behind the future head, and this region narrows during development while simultaneously elongating in the head-to-rump (anterior–posterior) direction (Fig. 1a, overleaf)¹. These movements, called convergence and extension, or convergent extension, occur when the cells become redistributed by converging on the embryo's midline through a process called

intercalation, forming a narrower but longer tissue². The cells intercalate by becoming polarized transversely in a bipolar manner, producing mobile protrusions called lamellipodia on their right and left surfaces that pull the cells between one another.

On page 364 of this issue, Ninomiya, Elinson and Winklbauer³ offer insights into how this all comes about. They show in frogs that the patterning events that occur along the anterior–posterior axis also polarize cell-intercalation activity and thereby, at a stroke, align trunk extension with anterior–posterior polarity.

Although much is known about the patterning of the head, trunk and tail tissues

along the anterior–posterior axis⁴, regulation of the convergent–extension movements that shape these tissues is less well understood⁵. One important unknown is how the local, polarized cell behaviour underlying these movements is globally aligned with the anterior–posterior array of tissues such that convergent extension pushes the head away from the rump.

Ninomiya *et al.*³ have addressed this problem. They first show that cells have an anterior–posterior positional identity that regulates their position along the anterior–posterior axis. When aggregates of anterior and posterior cells, labelled to keep track of their origin, were dissociated, mixed together and reaggregated, the cells recognized one another by their anterior–posterior level of origin, and sorted themselves into the correct position. Second, the authors found that the cells seem to use these position-related differences to switch on cell intercalation. In these experiments, intercalation occurred only in regions of contact between tissues originating at different anterior and posterior levels, not between tissues of the same level of origin (Fig. 1b). Significantly, extension occurred along the same axis as the mismatch of anterior–posterior levels.

Finally, Ninomiya *et al.*³ show that these position-dependent activities are related to the expression of two genes — called *Xenopus Brachyury* (*Xbra*) and *chordin* — that also regulate the anterior–posterior fate of cells. These genes are expressed in trunk tissues in 'countergradients' — *Xbra* expression is low anteriorly and high posteriorly, and *chordin* expression has the reverse pattern (Fig. 1b). Knowing that increasing concentrations of the signalling molecule activin boost the expression of *chordin* and reduce that of *Xbra*, the authors investigated whether graded activin treatment could also induce the position-dependent, intercalation-triggering activities.

They found that convergent extension occurred between strongly and moderately induced cell aggregates and between moderately and weakly induced ones, but not between two strongly induced or two weakly induced aggregates. Again, extension occurred along the same axis as the mismatch between strongly induced (equivalent to anterior) and weakly induced (equivalent to posterior) tissues; this is also the axis of differential gene expression. Although activin is a useful protein for these experiments, in the embryo it is more likely that a related signalling molecule called nodal, and its inhibitor lefty, are involved in the graded anterior–posterior gene expression^{6,7}, and perhaps in the graded properties that trigger intercalation.

All in all, these results are a major advance, but inevitably pose questions. What are the anterior–posteriorly graded cell properties? They might reflect differ-



100 YEARS AGO

The third annual meeting of the general committee of the Cancer Research Fund was held last Friday, July 8, at Marlborough House, the Prince of Wales occupying the chair. The report of the superintendent (Dr. Bashford) details the work that has been carried out during the past year. Specimens of new growths have been examined from a variety of animals, including fish and a wild mouse, showing that cancer occurs in animals in a wild state. Certain cells of malignant new growths have been found to present nuclear changes similar to those by which the sexual cells are prepared for fertilisation, and the fusion of nuclei has been demonstrated in tumours of the mouse. These observations suggest that the new growth of cancer is a mass of cells that has taken on an independent existence. Statistical investigations have also been carried out, and among other things do not support the widely spread belief that cancer is on the increase.

From *Nature* 14 July 1904.

50 YEARS AGO

Hydrogen in the manufacture of steel.

More than eighty years ago Thomas Graham, who was then Master of the Mint, observed that hydrogen forms a substantial proportion of the gases which are evolved when ferrous materials are heated in vacuum. It was not, however, until some fifteen or twenty years ago that the great importance of this gas began to be fully realized. In the solidification of the ingot, and right through to the final user, hydrogen can exert a distinctly detrimental effect, of which the steelmaker is now fully aware and takes proper precautions to avoid. Blow-holes in steel ingots, hair-cracks in forgings, and blisters in thin sheets are examples of the dangers which may arise from excessively high hydrogen contents, while the ductility of the material as measured by the reduction of area falls progressively as the hydrogen content rises. In the Review issued by Murex, Ltd., K. C. Barraclough provides a very useful, critical summary of existing knowledge regarding the effect of hydrogen on steel, in which a substantial amount of original data is also incorporated. The determination of the hydrogen content is by no means an easy matter in view of the fact only rarely does the amount of this gas in steel exceed 0.001 per cent.

From *Nature* 17 July 1954.

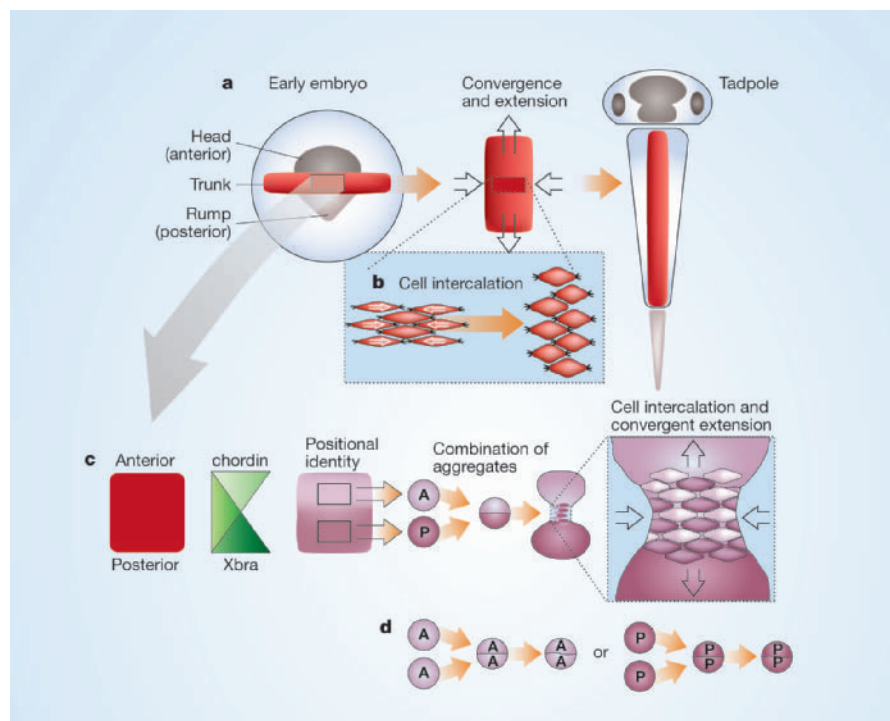


Figure 1 Separating the head from the rump. **a**, Chordamesodermal cells in the early embryo (red) undergo dramatic narrowing (convergence) and lengthening (extension) movements parallel to the anterior–posterior (head-to-rump) axis, thereby pushing the future head away from the rump or tail. **b**, The chordamesodermal cells intercalate transversely to form a narrower, longer array. The cells first become bipolar by developing large lamellipodia at their transverse ends. These exert traction, and as a result the cells become elongated, aligned with one another, and transverse with respect to the anterior–posterior axis. They finally intercalate to form a longer, narrower array. **c**, Ninomiya *et al.*³ show that the anterior–posterior axis has countergradients of *chordin* and *Xbra* gene expression associated with graded tissue properties that specify local positional identity. Aggregates of anterior (A) and posterior (P) tissues, in various combinations, show that intercalation and convergent extension occur in the region of contact between tissues originating at different anterior–posterior levels, but not (**d**) between those originating at the same level.

ences in cell–cell adhesion, which could explain the position-specific sorting in the cell-mixing experiments and the maintenance of anterior–posterior cell position in the embryos. How do cells measure these local differences, and what then triggers their transverse movement?

There is also the question of the scale on which they act. The authors' experiments suggest that, among intimately mixed cells, local differences in positional identity are sufficient to trigger cell sorting but not polarization, which seems to require a broader front between cells of differing positional identity. Also, the fact that evaluation of properties along the anterior–posterior axis triggers cell polarization suggests that the primary event here occurs at the anterior and posterior cell surfaces, rather than at the transverse surfaces. Perhaps lamellipodium formation is suppressed at these former surfaces, leaving it to occur transversely.

These findings in frogs suggest that convergent extension may be similar in other organisms. For instance, the 'germ band' of the fruitfly embryo converges and extends by transverse cell intercalation, and this process also elongates its body axis. The

intercalating germ-band cells are bipolar, and their activity relies on anterior–posterior or axial differences⁸, although in this case the differences are linked to so-called pair-rule genes, which are also involved in regulating anterior–posterior patterning. The genes are expressed in an alternating series of discrete transverse stripes instead of showing the graded differences in gene expression seen in frogs. But the important similarity is that, in both frog and fly, a comparison of anterior–posterior properties, linked to the expression of anterior–posterior patterning genes, triggers the polarization that underlies cell intercalation.

Ray Keller is in the Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22903, USA.

e-mail: rek3k@virginia.edu

- Vogt, W. *Wilhelm Roux Arch. EntwMech. Org.* **120**, 384–706 (1929).
- Shih, J. & Keller, R. *Development* **116**, 915–930 (1992).
- Ninomiya, H., Elinson, R. P. & Winklbauer, R. *Nature* **430**, 364–367 (2004).
- Harland, R. & Gerhart, J. *Annu. Rev. Cell Dev.* **13**, 611–667 (1997).
- Keller, R. *Science* **298**, 1950–1954 (2000).
- Branford, W. W. & Yost, H. J. *Curr. Biol.* **12**, 2136–2141 (2002).
- Dougan, R. M. *et al. Development* **130**, 1837–1851 (2003).
- Zallen, J. & Wieschaus, E. *Dev. Cell* **6**, 343–355 (2004).

Oceanography

Planktonic life on an ocean wave

Geophys. Res. Lett. **31**, L13301
doi:10.1029/2004GL019738 (2004)

A wave doesn't sound like a terribly surprising thing to find in the ocean, but the one reported by Meric A. Srokosz and colleagues is. It is a wave of plankton growth, and it propagates for three months over a distance of about 2,500 km before dying out. The plankton wave shows up in remote sensing of chlorophyll in the Indian Ocean in early 1999; similar waves are apparent in data from 2000 and 2002.

The wave moves from west to east — against the mean ocean-surface flow, the direction of large-scale ocean waves (Rossby waves) and the direction of eddy propagation. So the pulse of plankton blooming isn't caused by advection. Srokosz *et al.* believe it may be a reaction–diffusion wave, like those seen in oscillating chemical reactions and other spontaneously pattern-forming systems. 'Reaction' here means phytoplankton growth, and 'diffusion' is the mixing caused by turbulent eddies. A rough estimate of the wave propagation speed in such a reaction–diffusion process matches that seen experimentally (about 10 km per day).

Philip Ball

Chemistry

The grasping claws of DNA

Angew. Chem. Int. Edn **43**, 3550–3553 (2004)

A 'DNA machine' that can precisely control the concentration of the human blood-clotting factor, α -thrombin, does so by a novel grab-and-release mechanism, according to Wendy U. Dittmer and colleagues.

The machine is based on a known 15-base DNA sequence that adopts a chair-shaped conformation in the presence of potassium ions, which makes it bind strongly to α -thrombin. The authors modified this sequence with a 'switch' — an extra sequence of 12 bases that controls the binding ability of the entire strand. Under normal conditions, this switch does not affect the α -thrombin-binding ability of the strand. But a second DNA sequence can trigger the switch by binding to it; when added to the mixture, it destroys the machine's conformation and releases α -thrombin.

The molecular trigger can be removed from the mix by adding a DNA sequence that is fully complementary to it. This locks away the trigger, freeing up the switch and allowing the α -thrombin to be bound again. The DNA machine can bind and release α -thrombin through many cycles, and each step has a half-life of a few minutes.

Mark Peplow



Solid-state physics

Origin of electrical hum

Phys. Rev. Lett. **92**, 257202 (2004)

In the dead of night, our homes hum. The buzz emitted by domestic electrical appliances is the sound of a transformer flexing to the rhythm of the a.c. power supply, as magnetostriction creates tiny changes in volume. This effect has been poorly understood at a fundamental level, but now Th. Strässle *et al.* have identified its quantum origin.

An interaction called exchange coupling between the atoms of magnetic systems results from the spin of their electrons. This modifies the direct interaction via magnetic dipoles, so that in a periodically changing magnetic field the equilibrium distance between atoms also changes. In effect, the increase in elastic energy is compensated for by a reduction in magnetic energy. But it hasn't been clear whether this 'exchange striction' can really account for such volume changes.

Strässle *et al.* have measured the exchange energy by using inelastic neutron scattering to monitor the energy spectra of dimers of manganese ions at different pressures in the salt $\text{CsMn}_{0.28}\text{Mg}_{0.72}\text{Br}_3$, in which chains of Mg ions are punctuated by pairs of Mn ions. They measure the corresponding Mn–Mn distances using neutron diffraction. The observed exchange striction energy seems sufficient by itself to account for the changes in Mn–Mn excitation energy caused by squeezing.

Philip Ball

Huntington's disease

Complex difficulties

Cell **118**, 127–138 (2004)

Huntington's disease, a fatal adult condition characterized by uncontrolled movements and dementia, is linked to a defect in the huntingtin protein that means the molecule contains repeated glutamine sequences. This abnormality results in dysfunction and death among neurons in the brain.

Using mouse cells, Laurent R. Gauthier and colleagues have tried to figure out the molecular causes of this process. They find that undamaged huntingtin enhances the transport of brain-derived neurotrophic factor (BDNF) along microtubules inside neurons. Without this factor, neurons in regions such as the striatum, which sustains the most damage in Huntington's disease, die off. The glutamine-repeat defect renders huntingtin unable to perform this function, probably because it cannot form a proper complex with the motor proteins that power microtubule transport.

What's more, the administration of wild-type huntingtin restores microtubule transport of BDNF, the authors report. This hints at a possible strategy to rescue brain cells if the disease can be diagnosed before too much damage occurs.

Michael Hopkin

Developmental biology

From lung to intestine

J. Biol. **3**, 11 (2004)

In the metaplasias that sometimes portend cancer, cells in one part of the body take on the characteristics of those from another. Tadashi Okubo and Brigid L. M. Hogan have unearthed one of signals that convert lung cells into intestinal ones.

Okubo and Hogan first showed that, from around halfway through mouse gestation, budding lung tissues switch on the Wnt signalling pathway, which is implicated in numerous steps of organ growth and development. They then genetically engineered mouse embryos so that the level of Wnt signalling was boosted above normal in the growing lungs.

The transgenic embryos did not survive to birth and their lungs were missing many of the normal cell types. A survey of genes expressed in the organ showed that many genes normally active in pulmonary cells were suppressed and replaced by those normally active in cell types lining the intestine.

The result may explain the origin of particular forms of metaplasia in humans, such as premalignant stomach cancer or Barrett's oesophagus, when intestinal cells sprout out of place. The authors suggest that inflammation or injury might trigger stem cells to divide and to aberrantly switch on the Wnt pathway.

Helan Pearson

Pitch discrimination in the early blind

People blinded in infancy have sharper listening skills than those who lost their sight later.

Do blind people develop superior abilities in auditory perception to compensate for their lack of vision? They are known to be better than sighted people at orientating themselves by sound, but it is not clear whether this enhanced awareness extends to other auditory domains, such as listening to music or to voices. Here we show that blind people are better than sighted controls at judging the direction of pitch change between sounds, even when the speed of change is ten times faster than that perceived by the controls — but only if they became blind at an early age. The younger the onset of blindness, the better is the performance, which is in line with cerebral plasticity being optimal during the early years.

Auditory spatial localization is enhanced in early-blind subjects^{1–3}, but the effect of blindness on performance in the non-spatial auditory domain is less clear: blind subjects seem to be better at some tasks, such as the perception of chords, than others^{4–7}. Age at the onset of blindness, which could be an important factor, has not generally been considered.

We compared the performance of a group of early-blind subjects ($n=7$; ages, 21–40 years; onset of blindness, 0–2 years after birth), a group of late-blind subjects ($n=7$; ages, 24–46; onset of blindness, 5–45 years) and a group of sighted controls ($n=12$; ages, 21–37) in judging the direction of pitch change (for details, see supplementary information). In the task, which is based on psychophysical testing of sighted individuals^{8,9}, subjects hear two pure tones of different frequencies at each trial and have to decide whether the pitch is rising (second sound with higher pitch) or falling. In the reference condition (ST in Fig. 1), the pitch difference was one-eighth of an octave (1.5 semitones, or 150 cents) and the duration of each tone was 333 milliseconds (Fig. 1a).

Task difficulty was parametrically manipulated in both the temporal and spectral domains, either by successively dividing tone duration by two (temporal series; T6, 167 ms, up to T48, 20.8 ms) or by dividing the frequency spacing between the tones by two (spectral series; S16, 1/16 of an octave, up to S128, 1/128 of an octave or 4.7 cents). Eight different tone pairs (one rising and one falling, at each of four different frequencies; frequency range, 500–1,000 hertz) were used for each difficulty level. Subjects heard stimuli through headphones binaurally and reported the pitch-change direction by pressing a key.

For all groups, performance was best in the reference condition (ST) and was significantly reduced when either the tone duration (Fig. 1b; 'T' bars) or the frequency difference

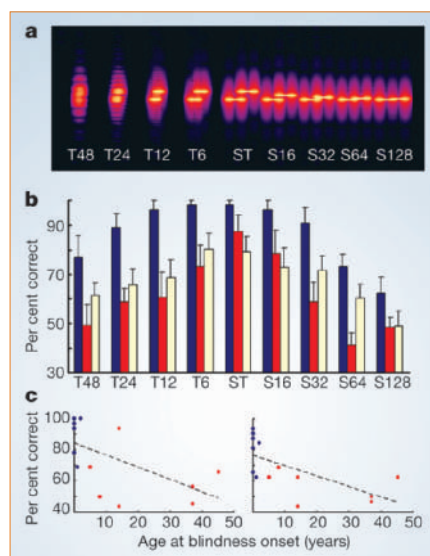


Figure 1 Judgement of direction of pitch change by early-blind, late-blind and sighted participants. **a**, Spectrogram of examples of the pure tones used in the task for temporal (T) and spectral (S) series; time increases horizontally and frequency increases vertically; colour intensity reflects the energy of the stimuli. Each tone pair shows a representative example of stimuli used in each of the nine conditions, chosen here with a rising pitch from the first to the second tone. The two tones in each pair were always presented in succession; the impression of temporal overlap for the most rapid conditions (temporal series) is an artefact related to the influence of the acoustic uncertainty principle on the time–frequency representation. **b**, Proportion of correct answers ($\pm$ s.e.m.) given by early-blind (blue bars), late-blind (red bars) and sighted (white bars) subjects for the nine conditions. **c**, Significant negative correlation between age of blindness onset and behavioural performance in the pitch-direction task in temporal (left; T6 to T48) and spectral (right; S16 to S128) conditions. Blue and red dots represent early- and late-blind groups, respectively.

between the tones (Fig. 1b; 'S' bars) was decreased ($P<0.001$). Early-blind subjects showed significantly better overall performance than either the late-blind or sighted subjects (Fig. 1b; main effect, $P=0.008$; post hoc Tukey: $P=0.009$ and $P=0.026$, respectively). The results followed this pattern for both the temporal ($P=0.011$) and the spectral ($P=0.011$) parts of the data set. Performance by the early-blind group for the most rapid condition (T48) was equivalent to that of sighted subjects in the easiest condition (ST), which comprised durations that were greater by an order of magnitude. No significant difference could be detected between late-blind and sighted subjects ($P=0.67$).

There was a significant negative correlation between age of blindness onset and overall performance (partial correlation $r=-0.65$, $P=0.012$) across the entire group of blind subjects; this relation held even after accounting for the duration of blindness

($r=-0.49$, $P=0.045$, one-tailed). Again, this finding held for both the spectral (Fig. 1c, right; $r=-0.65$, $P=0.011$) and the temporal (Fig. 1c, left; $r=-0.60$, $P=0.024$) parts of the data set.

Early-blind subjects were better than both late-blind and sighted subjects at determining the direction of pitch change, for different temporal as well as spectral levels. We conclude that compensatory auditory mechanisms following visual deprivation must extend beyond the spatial domain. Our finding that a large part of the variance (42%) could be accounted for by the age of blindness onset may explain why conflicting results are found when early- and late-blind subjects are pooled together^{4,6}. Moreover, it is in agreement with the idea that cerebral plasticity is more efficient at early developmental stages.

Frédéric Gougoux*, Franco Lepore*, Maryse Lassonde*, Patrice Voss*, Robert J. Zatorre†, Pascal Belin*

*Centre de Recherche en Neuropsychologie et Cognition, Département de Psychologie, Université de Montréal, CP 6128, Succursale Centre-ville, Montréal, Québec H3C 3J7, Canada
e-mail: pascal.belin@umontreal.ca

†Montreal Neurological Institute, McGill University, Montréal, Québec H3A 2B4, Canada

1. Röder, B. et al. *Nature* **400**, 162–166 (1999).
2. Lessard, N., Paré, M., Lassonde, M. & Lepore, F. *Nature* **395**, 278–280 (1998).
3. Muchnick, C., Efrati, M., Nemeth, E., Malin, M. & Hildesheimer, M. *Scand. Audiol.* **20**, 19–23 (1991).
4. Sakurabayashi, H., Sato, Y. & Uehara, E. *J. Psychol. Blind* **1**, 3–10 (1956).
5. Pitman, D. J. *Am. Found. Blind. Res. Bull.* **11**, 63–79 (1965).
6. Juurma, J. *Am. Found. Blind. Res. Bull.* **14**, 109–122 (1967).
7. Stankov, L. & Spilbury, G. *Appl. Psychol. Meas.* **2**, 491–503 (1978).
8. Johnsrude, I. S., Penhune, V. B. & Zatorre, R. J. *Brain* **123**, 155–163 (2000).
9. Zatorre, R. J. & Belin, P. *Cereb. Cortex* **11**, 946–953 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Fisheries

Mislabelling of a depleted reef fish

Any fish species that appears to be readily available in the marketplace will create an impression among the public that there is a plentiful supply of that fish in the sea, but this may belie the true state of the fisheries' stock. Here we use molecular genetic analysis to show that some three-quarters of the fish sold in the United States as 'red snapper' — the US Food and Drug Administration's legally designated common name for *Lutjanus campechanus*¹ — belong to another species. Mislabelling to this extent not only

defrauds consumers but could also adversely affect estimates of stock size if it influences the reporting of catch data that are used in fisheries management.

The red snapper, or *L. campechanus*, is found in offshore waters around coral reefs and rocky outcroppings² and is one of the most economically important fisheries in the Gulf of Mexico, with greater total landings and commanding higher prices than any other snapper species^{3,4}. In 1996, the Gulf of Mexico Fishery Management Council and the United States Department of Commerce declared that *L. campechanus* was grossly overfished and called for strict management measures to restore stocks to sustainable levels.

Tight restrictions may create an economic incentive for seafood substitution, whereby less valuable species are mislabelled and sold under the names of more expensive ones. Substitutions among closely related fish species are difficult to detect, because most distinguishing features are lost during processing. Although the name 'red snapper' is applied to other species in different countries, only *L. campechanus* can legally be labelled as red snapper in the United States¹.

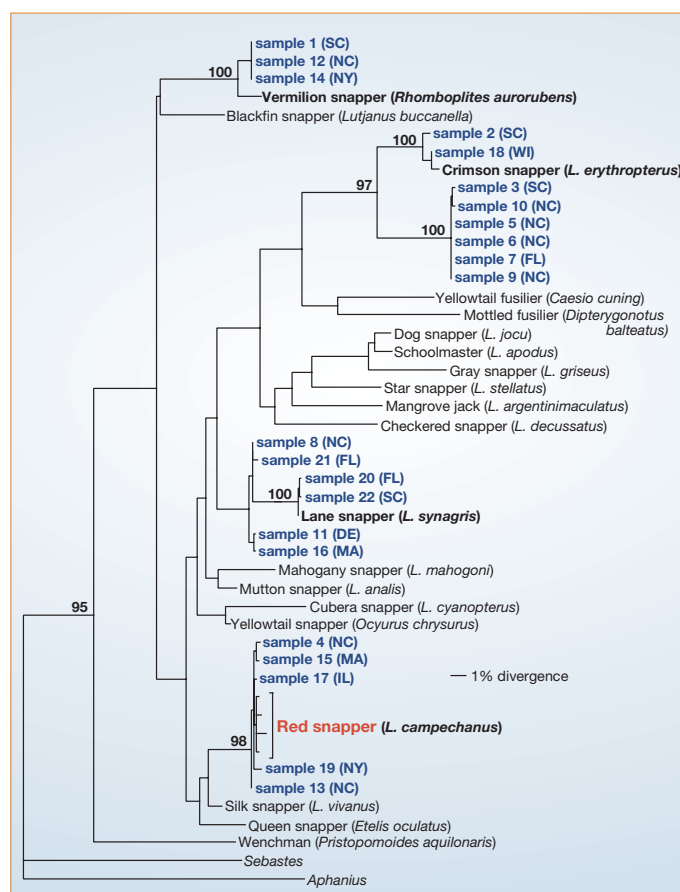
To investigate the extent of possible mislabelling in the United States, we did a molecular phylogenetic analysis of fish purchased from nine vendors in eight states (Fig. 1). We compared DNA sequences from retail samples with those deposited in GenBank, which contains sequences from all snappers for which landing data are recorded by the US National Marine Fisheries Service (NMFS).

Our results show that mislabelling is a large-scale phenomenon: 77% of fish sold as red snapper (17/22) were in fact other species (Fig. 1). This estimate has a margin of error of 17%, meaning that between 60% and 94% of fish sold as red snapper in the United States are mislabelled. Among the sequences from fish that are not *L. campechanus*, five are grouped very closely with those from other Atlantic species. Of these, sequences from two are nearly identical to those from the lane snapper, differing only by single-nucleotide substitutions (Fig. 1). The three other putative Atlantic snappers in our sample are grouped with vermillion snapper, with only about 2% sequence divergence.

Surprisingly, more than half of our analysed sequences are either grouped closely with species from other regions of the world or represent species too rare to have yet been included in molecular phylogenetic surveys⁵. For example, two fish can be definitively identified as crimson snappers, a species from the Indo-West Pacific. And sequences from several other fish form clades that do not seem to be closely related to any reference sequences. These fish represent species that are either not caught in US waters or not yet individually managed by the NMFS.

Mislabelling of fish can occur at different points in the commercial process: for

Figure 1 Maximum-likelihood tree (GTR + I + G model) of cytochrome-*b* DNA sequences from retail 'red snapper' (GenBank accession numbers: AY294187–205; AY651957–59) and reference sequences (AF23–9677–78, 80–82; AF240–750; AF381270; AF031516; AF299290; U26949, 51–58, 61–62). Numbers at nodes are bootstrap percentages (1,000 replicates) from an unweighted parsimony search. Sequences of 953 base pairs were obtained using the polymerase chain reaction and the primers CB12F (5'-TGGCAAGCTAC-GCAAAC-3') and CB13R (5'-TATTCGCGGATTCAGGT-AA-3') followed by automated sequencing on an ABI 377. Phylogenetic analyses were completed using PAUP* 4.0 software¹³. State abbreviations: DE, Delaware; FL, Florida; IL, Illinois; MA, Massachusetts; NY, New York; NC, North Carolina; SC, South Carolina; WI, Wisconsin.



example, retail mislabelling probably explains the presence of crimson snapper in our sample, because these fish occur in an entirely different part of the world from *L. campechanus*. Fish substitutions may occur on the boat or at the dock, because species that are morphologically similar to *L. campechanus* are caught together, as is the case for most reef fisheries⁶. The implications for resource management are important: substitutions made before commercial landing data are submitted to the regulatory agencies will cause inflation of catch estimates for the most desirable species and simultaneous underreporting of catches for less desirable, and potentially unmanaged, species. Although it is widely acknowledged that intentional misreporting and unintentional misidentification of species can both bias fishery-dependent data^{7–11}, the contribution of this effect cannot be evaluated unless fish are sampled before the collection of catch data.

The remarkable extent of product mislabelling in the case of *L. campechanus* threatens to distort the status of fish stocks as perceived by consumers, which contributes to the false impression that the supply of fish is keeping up with demand¹².

Peter B. Marko*, **Sarah C. Lee***, **Amber M. Rice†**, **Joel M. Gramling†**, **Tara M. Fitzhenry†**, **Justin S. McAlister†**, **George R. Harpert†**, **Amy L. Moran***
Departments of *Marine Sciences and †Biology,

University of North Carolina, Chapel Hill, North Carolina 27599–3300, USA

e-mail: pmarko@unc.edu

- Office of Seafood, Food and Drug Administration (<http://www.cfsan.fda.gov/~frf/seaintro.html>).
- Moran, D. Biological report 82(11.83) TR EL-82-4 (National Wetlands Research Center, Louisiana, 1988).
- Huang, T.-S., Marshall, M. R., Kao, K.-J., Ottwell, W. S. & Wei, C.-I. *J. Agric. Food Chem.* **43**, 2301–2307 (1995).
- National Marine Fisheries Service, Fisheries Statistics and Economics Division, Silver Spring, Maryland (<http://www.st.nmfs.gov/>).
- Salver, S. K., Freshwater, D. W. & Walsh, P. J. *Copeia* **3**, 715–719 (1996).
- Johannes, R. E. *Trends Ecol. Evol.* **13**, 243–246 (1998).
- Watson, R. & Pauly, D. *Nature* **414**, 689–695 (2001).
- Reef Fish Assessment Panel Report (Gulf of Mexico Fishery Management Council, 1994).
- Marine Fisheries Stock Assessment Improvement Plan (National Marine Fisheries Service, 2001).
- Atlantic Billfish Research Plan (National Marine Fisheries Service, 2002).
- Porch, C. E. & Cass-Calay, S. L. *Status of the Vermilion Snapper Fishery in the Gulf of Mexico: Assessment 5.0 SFD-01/02-129* (Southeast Fisheries Science Center, Miami, Florida, 2001).
- Review of Existing Public Opinion Data on the Oceans (The Ocean Project, 1999).
- Swofford, D. L. PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4 (Sinauer, Sunderland, Massachusetts, 2001).

Competing financial interests: declared none.

brieF communications arising online

▶ www.nature.com/bca

Ecology: Climate-change effect on Lake Tanganyika?

W. W. Eschenbach (doi:10.1038/nature02689)

Reply: C. M. O'Reilly, P.-D. Plisnier, A. S. Cohen, S. R. Alin (doi:10.1038/nature02737)

Ecology

Climate-change effect on Lake Tanganyika?

Arising from: C. M. O'Reilly, S. R. Alin, P.-D. Plisnier, A. S. Cohen & B. A. McKee *Nature* **424**, 766–768 (2003)

In their analysis of Lake Tanganyika's ecosystem¹, O'Reilly *et al.* claim that climate change, in the form of rising temperatures and falling winds, is causing a decline in the lake's productivity. However, their own data show that air temperatures were either steady or dropped slightly between 1952 and 1978, rising only slightly between 1980 and 1992, and that wind speeds have increased by 35% since 1985. These climate changes therefore have no correlation with either lake temperature or productivity, so it cannot be inferred from their data that climate change is the cause of the productivity decline.

The authors claim that trend lines indicate that air temperatures at the lake warmed by about 0.5–0.7 °C. However, trend lines do not reveal changes over time — gaussian averaging gives a better insight. The authors' data¹ for Mbala and Bujumbura (Fig. 1) reveal that there was no change or a slight decrease in average temperature between 1952 and 1975; Bujumbura wind speed was unchanged between 1967 and 1975, and Mbala wind speed increased over that period.

If climate change were the reason for the decline in productivity, we should see no such decline between 1952 and 1975, or might even see an increase due to rising wind speeds at Mbala. Nor should we see any change in water temperature, as there was no air-temperature change. However, despite steady temperatures and rising winds, the authors' values for productivity (Fig. 3 of ref. 1) show a steady decline during that time, and their water temperature (Fig. 2a of ref. 1) shows an increase of roughly 0.25 °C.

After about 1975, both the temperature and wind records become suspect, with a considerable amount of missing data. There is an abrupt 0.7 °C rise in the temperature in Mbala in 1980, accompanied by a large reduction in standard deviation (see the

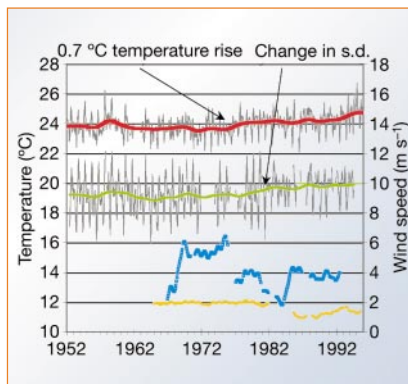


Figure 1 Historical monthly weather records for Lake Tanganyika. Temperature data: red line, from Bujumbura; green line, from Mbala (gaussian average over six years); grey lines, monthly temperatures. Wind-speed data: blue squares, from Mbala; yellow dots, from Bujumbura (gaussian average over two years). Created from the original weather data of O'Reilly *et al.*¹.

authors' Fig. 1a: standard deviation 1952–80, 1.22; s.d. 1980–94, 0.95). This looks as though there has been either an artificial splicing of two records or a change in thermometer location, particularly as the same pattern is not present in Bujumbura.

The later parts of both wind records are particularly poor. After a year's gap in 1976, the Mbala wind records resume sporadically (a full quarter of the post-1976 Mbala monthly data is missing), with much lower values and standard deviation (s.d. 1967–75, 2.0; s.d. 1976–92, 1.4).

Between 1967 and 1982, the Bujumbura wind data are complete; there is not a single missing month. But after a three-year hiatus (1982–85), 17% of the Bujumbura monthly data is missing, and radically lower values are reported. I suggest that the wind figures after the gaps should be used with caution unless they can be confirmed by other observations.

For the sake of argument, let us assume

that the wind records of O'Reilly *et al.* are correct. From their data¹, we can draw several conclusions. First, air temperatures were steady or dropped slightly at Lake Tanganyika between 1950 and about 1978. They rose abruptly, and then rose only slightly during 1980–90, after which only temperatures at Bujumbura rose. However, we see no sign of this pattern in either water temperature or lake productivity (Figs 2, 3 of ref. 1). In particular, water temperatures rose between 1950 and 1978, whereas air temperatures were unchanged or dropped slightly; water temperatures fell over 1993–96 (and continued to fall through 2003), whereas air temperatures rose slightly from 1993 to the end of the record in 1996. Also, all their productivity indices dropped between 1952 and 1965, when air temperatures were dropping slightly rather than rising.

Second, the wind-speed data of O'Reilly *et al.* have several problems, but purport to show a rise in Mbala wind speed between 1967 and 1975, a sudden drop in 1976 and a dip in 1978–84, followed by a 35% increase that lasted until 1996. The authors say that wind in Mbala is critical for productivity, but their productivity data do not reflect these wind changes. The productivity data series that extends past 1970 shows an overall productivity decrease from 1970 to 1996 (one series shows a short-lived rise in about 1990).

Third, as data for both air temperature and wind are so poor, and as there is no correlation between the reported temperature changes and wind changes on the one hand and productivity and water-temperature data on the other, the null hypothesis (that changes in productivity are not the result of climate change) cannot be rejected. Based on the authors' data, therefore, the productivity changes cannot be ascribed to climate change.

Willis W. Eschenbach

PO Box 351, Pacific Harbour, Fiji, South Pacific

e-mail: willis@taunovobay.com

doi:10.1038/nature02689

1. O'Reilly, C. M., Alin, S. R., Plisnier, P.-D., Cohen, A. S. & McKee, B. A. *Nature* **424**, 766–768 (2003).

O'Reilly *et al.* reply — Eschenbach¹ questions the validity of our climate record and our proposed link between climate and lake ecosystem change. Here we address the three main issues: our meteorological data, the timing of limnological changes, and the warming of Lake Tanganyika.

Our meteorological data are sufficient to indicate likely trends in local climate for three reasons.

First, there has been no change in either the location of the station or the instrumentation in Bujumbura, and an independent analysis of the data shows a 0.6 °C warming over the past 80 years (M. Shiramanga, per-

sonal communication). The Mbala meteorological station has not changed location and underwent a change in instrumentation for temperature (but not for wind speed) in 1984 (I. Jere, personal communication). The heat-island effect is an unlikely cause of warming because there has been no urban development at either location.

Second, the pattern of reduced variability seen in the latter part of the air-temperature record is consistent with reduced diurnal temperature ranges (as much as 0.5–1.0 °C since the 1950s; ref. 2) observed elsewhere in eastern Africa and around the world^{2–4}. A change in standard deviation of the climate

record does not constitute proof that the jump in temperature and/or wind speed is not real. Gizozi and Mparambo weather stations in Burundi indicate a comparable rise in temperature of 0.2 °C in each decade, including a sharp increase in air temperature between 1977 and 1979 (ref. 5).

Third, although these local climate data lack the continuity and resolution (that is, sub-monthly) required for extensive statistical analysis, the magnitude and patterns are consistent with longer and higher-resolution regional records^{2,3,6–10}.

Limnological changes would not be expected to coincide directly in time with

climate shifts or transitions. This is because the climate, limnological and core data derive from different locations in this large ecosystem. Furthermore, both limnological records and core data have low resolution relative to the monthly climate data, making it difficult to determine the exact timing of changes in the lake. And finally, there are time lags associated with the lake's response to warming, which may be magnified as one travels up the foodweb.

Lake Tanganyika has warmed. In our view, there is no plausible mechanism other than climate change to explain this warming adequately. Even if there had been no change in wind speed, 97% more energy would be required to mix the lake on the basis of the increased temperature stratification. There-

fore, unless wind speeds increased dramatically, a reduction in deep-water nutrient input would still be expected, with a subsequent decrease in primary productivity.

Although the available meteorological time-series data are not optimal, the uncontested warming of the lake and a multitude of regional climate observations confirm that climate is the most parsimonious explanation for the recent productivity changes in Lake Tanganyika.

**C. M. O'Reilly*, P.-D. Plisnier,
A. S. Cohen, S. R. Alin**

**Environmental Science Program, Vassar College,
Poughkeepsie, New York 12604, USA*

e-mail: caoreilly@vassar.edu

doi:10.1038/nature02737

1. Eschenbach, W. W. *Nature* doi:10.1038/nature02689 (2004).

2. Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. *Clim. Res.* **17**, 145–168 (2001).
3. King'uyu, S. M., Ogallo, L. A. & Anyamba, E. K. *J. Clim.* **13**, 2876–2886 (2000).
4. Nicholls, N. et al. in *Climate Change 1995: The Science of Climate Change* (eds Houghton, J. T., Filho, L. G. M., Callendar, B. A., Kattenburg, A. & Maskell, K.) 133–192 (Cambridge University Press, Cambridge, 1996).
5. *Ministère de l'Aménagement du Territoire et de l'Environnement Convention Cadre des Nations Unies sur les Changements Climatiques: Première Communication Nationale* (Burundi, 2001).
6. Mears, C. A., Schabel, M. C. & Wentz, F. J. *J. Clim.* **16**, 3650–3664 (2003).
7. Houghton, J. T. et al. *Climate Change 2001: The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, 2001).
8. New, M., Hulme, M. & Jones, P. *J. Clim.* **13**, 2217–2238 (2000).
9. Epstein, P. et al. *Bull. Am. Meteorol. Soc.* **79**, 409–417 (1998).
10. Loewinson, M. E. *Lancet* **343**, 714–718 (1994).

The scientific impact of nations

What different countries get for their research spending.

David A. King

The ability to judge a nation's scientific standing is vital for the governments, businesses and trusts that must decide scientific priorities and funding. In this paper I analyse the output and outcomes from research investment over the past decade, to measure the quality of research on national scales and to set it in an international context. There are many ways to evaluate the quality of scientific research, but few have proved satisfactory. My analysis updates and extends the groundbreaking work by May¹, which covered 1981–94, and draws on a study of 1993–2002 commissioned by the UK Office of Science and Technology (OST)². Although the OST study's admittedly parochial objective is to evaluate the United Kingdom's performance in science and engineering research, this paper should be of more widespread interest as it provides metrics for judging achievement, and analyses of the output and outcomes of other countries.

To measure the quantity and quality of science in different nations, I have analysed the numbers of published research papers and reviews, and their citations. All data were provided by Thomson ISI, previously known as the Institute for Scientific Information, which indexes more than 8,000 journals in 36 languages, representing most significant material in science and engineering.

One potential problem with this type of bibliometric analysis is that individual papers can skew the results. For instance, a paper may be highly cited because it has been discredited, or because its authors over-cite their own work. However, the large number of papers in this study should smooth out such distortions. The heads of large research institutions may also be authors on papers to which their contribution was largely indirect. However, my analysis considers only authors' country of origin, that is where they are working, rather than individual names. Finally, citation analyses must not be used to compare different disciplines. For example, papers in medical research get many more citations than those in mathematics. An aggregate of citations across disciplines will therefore lead to high-citation disciplines swamping the low, an issue I will consider later.

I made comparisons across 31 countries (the comparator group) including the G8 group (italicized) and the 15 countries of the European Union before the 2004 accession (EU15). The countries are: Australia,

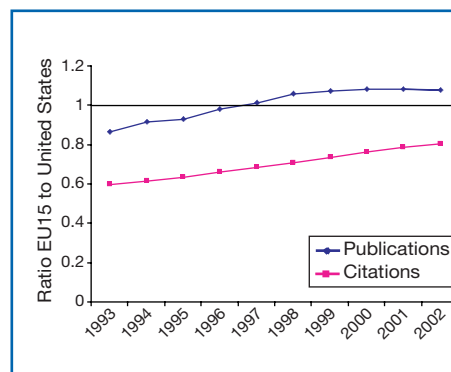


Figure 1 Comparing Europe with the United States. Ratio of the publications and citations of the 15 European Union countries in the comparator group (EU15) to the United States on ISI databases in 1993–2002. The EU15 total contains some duplication because of papers jointly authored between countries in the EU group. Counts for papers and citations are totals for country (or group) for the stated year.

Austria, Belgium, Brazil, *Canada*, China, Denmark, Finland, *France*, *Germany*, Greece, India, Iran, Ireland, Israel, *Italy*, *Japan*, Luxembourg, the Netherlands, Poland, Portugal, *Russia*, Singapore, Spain, South Africa, South Korea, Sweden, Switzerland, Taiwan, the *United Kingdom* and the *United States*. This group accounted for more than 98% of the world's highly cited papers, defined by Thomson ISI as the most cited 1% by field and year of publication. The world's remaining 162 countries contributed less than 2% in total.

The premier league

Table 1 shows the total publications and the citation analyses for the comparator group, covering all fields of science and engineering. Each cited paper is allocated once to every country in which an author is based, so some papers are counted twice or more. The sum of national publications exceeds the world total because of international collaboration. National shares similarly sum to more than 100%. The totals for grouped publications (EU15, world) are adjusted to take this into account, by removing duplicate papers with multiple national authorship, giving an accurate net total.

The rank order in Table 1 is only marginally affected by the variable selected and is similar to that for 1981–94 (ref. 1). It is not particularly sensitive to which of the two citation analyses are used. The percentage share of world citations (citations share index) is the national share of citations in all disciplines, and so gives more weighting to the more heavily cited disciplines, such as medicine. The United States easily heads the list of nations in the volume of publications and citations and the share of top 1% cited papers, although the EU15 countries now publish more papers than the United States (Fig. 1) and are not far behind on citations. The United Kingdom is second on the two citations listings, but Japan publishes

almost as many papers and Germany is closing the gap in citations.

The United States' share of citations dropped by about 3% between 1993 and 2002, whereas the United Kingdom's and Germany's have each increased. The corrected UK share was 10.87% in 1993–97, 11.39% in 1997–2001 and 11.6% in 2001. The nations with the most citations are pulling away from the rest of the world. The G8 countries are in this premier division, apart from Russia, which has seen a radical drop in science investment since the break-up of the Soviet Union. The smaller northern European countries are snapping at the top seven's heels, and would be in this premier division if combined.

Although national contribution to world science is clearly an important indicator, it is also useful to compare outputs and outcomes relative to population and gross domestic product (GDP). Figure 2 compares 'wealth intensity', or GDP per person, with 'citation intensity' — citations per unit GDP. The smaller nations in the group — the Scandinavian countries, Israel, the Netherlands and Switzerland — all perform strongly by this measure, with the latter in a significant lead. The nations with a wealth intensity between US\$20,600 (Spain) and US\$35,800 (the United States) per person cover a nearly tenfold spread in citation intensity, with little correlation in this range. Among the outliers from the G8 are the United States and Japan, both below the average citation intensity for the full comparator group, and the United Kingdom, which is above average. At the other end of the scale, although the GDPs of China and India place them, respectively, second and fourth in the world, each has a low wealth and citation intensity.

The number of citations per paper is a useful measure of the impact of a nation's output. The data can be re-based, to avoid distortions due to different citation rates in

different disciplines, by normalizing to the average for each field and accounting for year of publication. The figure obtained is the re-based impact (RBI). Table 2 shows the average RBI over all disciplines among the 31 nations for 1993–2002, and also, to demonstrate the trend, the figure for 2002. Of the G8 nations, the United States heads this table, but once again the gap between the United States, the United Kingdom and Germany has narrowed significantly over the study period. As the US RBI shows a small increase over this period, its drop in citation share (Table 1) is due to a smaller number of published papers, with a higher average quality.

It is unlikely that these results reflect an anglophone bias, as it is now accepted that all high-quality papers are published in English. But anecdotal evidence suggests that preferential US citing of US papers may distort the analyses, given the sheer size of the US contribution. It is possible that Japan and

Russia, being more scientifically isolated than the other major players, suffer particularly in this respect.

Dividing disciplines

Governments also need an indication of disciplinary strengths and weaknesses based on international comparisons. Comparing one discipline across different countries should be easier than comparing two disciplines within one country.

There is a variety of ways to categorize disciplines. The Organisation for Economic Co-operation and Development (OECD) database uses five broad groups: medical sciences; natural sciences; agricultural sciences; engineering and technology; and social sciences. In the United Kingdom, the assessment performed to determine public research funding is based on 68 disciplinary units. A bibliometric analysis³ of seven countries by Adams, including England but not the United Kingdom, based on these units

for 1988–96 grouped the 68 units into blocks. Finally, the OST's grant from the UK treasury is divided mainly among seven research councils, each run at arm's length from government.

Taking all of these maps into account, and also noting similarities of journal usage by UK researchers from within the 68 disciplinary units, our contractors based their analysis² on seven categories: clinical medicine; preclinical medicine and health; biological sciences; environment; mathematics; physical sciences; and engineering.

This national and disciplinary disaggregation process produces a substantial database², but it can be presented in a compressed form using footprint plots (also known as spider or radial plots). The citation share for each category gives a measure of research impact. The aggregated citation share for each nation is shown for comparison. The larger the national footprint, the bigger the impact on international science. The foot-

Table 1 Rank order of nations based on share of top 1% of highly cited publications, 1997–2001

Country	Publications				Citations				Top 1% highly cited publications			
	1993–97		1997–2001		1993–97		1997–2001		1993–1997		1997–2001	
	Total	Per cent world	Total	Per cent world	Total	Per cent world	Total	Per cent world	Total	Per cent comparator group	Total	Per cent comparator group
United States	1,248,733	37.46	1,265,808	34.86	21,664,121	52.3	10,850,549	49.43	22,710	65.6	23,723	62.76
EU15 (net total)	1,180,730	35.42	1,347,985	37.12	15,147,205	36.57	8,628,152	39.3	11,372	32.85	14,099	37.3
United Kingdom	309,683	9.29	342,535	9.43	4,502,052	10.87	2,500,035	11.39	3,853	11.13	4,831	12.78
Germany	268,393	8.05	318,286	8.76	3,575,143	8.63	2,199,617	10.02	2,974	8.59	3,932	10.4
Japan	289,751	8.69	336,858	9.28	3,123,966	7.54	1,852,271	8.44	2,086	6.03	2,609	6.9
France	203,814	6.11	232,058	6.39	2,638,563	6.37	1,513,090	6.89	2,096	6.05	2,591	6.85
Canada	168,331	5.05	166,216	4.58	2,315,140	5.59	1,164,450	5.3	2,002	5.78	2,195	5.81
Italy	122,398	3.67	147,023	4.05	1,535,208	3.71	964,164	4.39	1,151	3.32	1,630	4.31
Switzerland	57,664	1.73	66,761	1.84	1,113,886	2.69	647,013	2.95	1,196	3.45	1,557	4.12
Netherlands	83,600	2.51	92,526	2.55	1,335,748	3.22	759,027	3.46	1,111	3.21	1,435	3.8
Australia	89,557	2.69	103,300	2.84	1,078,746	2.6	623,636	2.84	852	2.46	1,049	2.78
Sweden	63,757	1.91	72,927	2.01	1,007,418	2.43	548,112	2.5	748	2.16	930	2.46
Spain	79,121	2.37	103,454	2.85	813,722	1.96	559,875	2.55	467	1.35	785	2.08
Belgium	40,147	1.2	48,010	1.32	574,095	1.39	339,895	1.55	482	1.39	639	1.69
Denmark	31,808	0.95	37,198	1.02	508,183	1.23	295,004	1.34	445	1.29	570	1.51
Israel	41,804	1.25	45,944	1.27	517,027	1.25	293,039	1.33	449	1.3	568	1.5
Russia	121,505	3.65	123,629	3.4	509,105	1.23	315,016	1.43	366	1.06	501	1.33
Finland	28,727	0.86	34,690	0.96	427,873	1.03	250,456	1.14	308	0.89	416	1.1
Austria	26,100	0.78	33,598	0.93	332,145	0.8	218,493	1	250	0.72	383	1.01
China	68,661	2.06	115,339	3.18	392,055	0.95	341,519	1.56	153	0.44	375	0.99
South Korea	26,838	0.81	55,739	1.53	183,122	0.44	192,346	0.88	97	0.28	294	0.78
Poland	34,680	1.04	42,852	1.18	237,622	0.57	155,310	0.71	170	0.49	231	0.61
India	72,877	2.19	77,201	2.13	316,461	0.76	188,481	0.86	112	0.32	205	0.54
Brazil	27,874	0.84	43,971	1.21	211,460	0.51	155,357	0.71	100	0.29	188	0.5
Taiwan	32,620	0.98	45,325	1.25	216,852	0.52	150,743	0.69	91	0.26	151	0.4
Rep. Ireland	9,880	0.3	12,779	0.35	104,442	0.25	75,893	0.35	86	0.25	196	0.36
Greece	16,463	0.49	22,333	0.62	128,646	0.31	89,822	0.41	76	0.22	113	0.3
Singapore	9,030	0.27	15,306	0.42	63,288	0.15	55,929	0.25	39	0.11	97	0.26
Portugal	8,102	0.24	13,583	0.37	74,196	0.18	62,814	0.29	43	0.12	96	0.25
South Africa	1,7461	0.52	18,123	0.5	121,598	0.29	67,916	0.31	51	0.15	81	0.21
Iran	2,152	0.06	4,813	0.13	10,706	0.03	12,325	0.06	5	0.01	14	0.04
Luxembourg	300	0.01	430	0.01	2,736	0.01	1,979	0.01	2	0.01	2	0.01
World (net total)	3,333,464	106.23	3,631,368	108.94	41,425,399	118.27	21,953,043	122.97	34,982	127.43	38,263	136.5

This part of the analysis uses a five-year publication window for all disciplines. For papers published 1993–97, the total accumulation of citations to the year 2002 is included. For papers published 1997–2001, the total number of citations to the year 2002 is also included but, given the shorter time period, fewer citations will have accumulated.

The main source of internationally comparable data on research funding, staff and training is the OECD (see 'statistics' at <http://www.sourceoecd.org/content/html/index.htm>). Data also come from the 2002 editions of the *Main Science and Technology Indicators* and *Basic Science and Technology Statistics*. Accuracy and reliability are discussed in ref. 2. The *Frascati Manual* data definitions and their interpretations of OECD data have been adhered to wherever feasible.

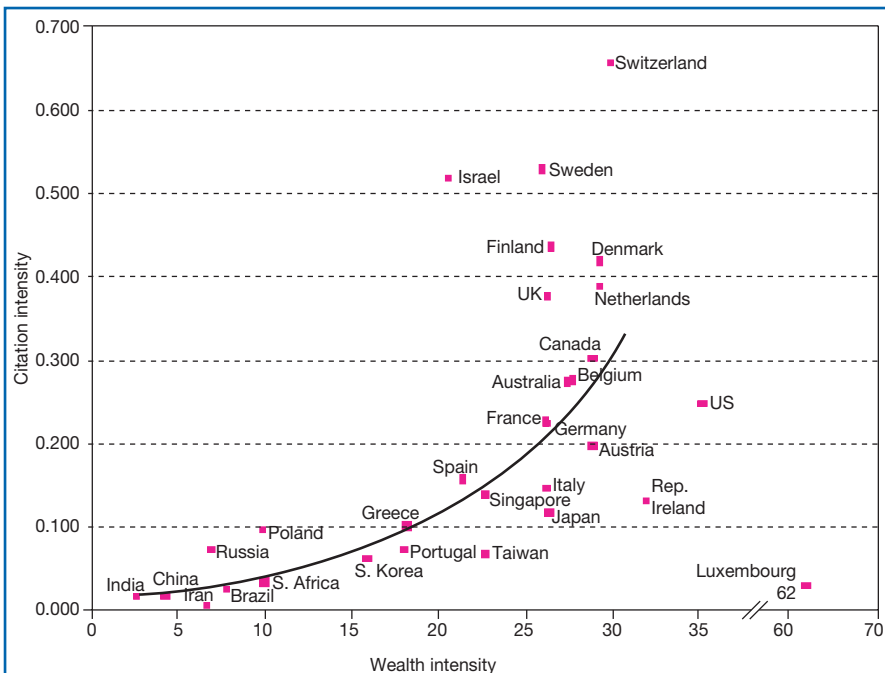


Figure 2 Comparing economic and scientific wealth. National science citation intensity, measured as the ratio of the citations to all papers to the national GDP, shown as a function of the national wealth intensity, or GDP per person, for the 31 nations in the comparator group. GDP and wealth intensity are given in thousands of US dollars at 1995 purchasing-power parity. Sources: Thomson ISI, OECD and the World Bank.

Table 2 Citation rate per paper normalized to show re-based impact (RBI)

Country	2002	1993–2002
Switzerland	1.70	1.59
United States	1.48	1.41
Denmark	1.48	1.33
United Kingdom	1.39	1.21
Netherlands	1.39	1.33
Germany	1.33	1.15
Austria	1.24	1.09
Belgium	1.21	1.17
Sweden	1.21	1.25
Canada	1.18	1.13
Finland	1.18	1.20
France	1.12	1.07
Italy	1.12	1.07
Australia	1.09	1.01
Israel	1.09	1.05
Rep. Ireland	1.03	0.93
Spain	0.97	0.89
Luxembourg	0.94	0.82
Japan	0.91	0.90
Portugal	0.82	0.80
Poland	0.76	0.61
South Africa	0.76	0.61
Greece	0.70	0.67
South Korea	0.64	0.61
Singapore	0.61	0.61
Brazil	0.58	0.62
Russia	0.55	0.40
China	0.55	0.51
Taiwan	0.55	0.56
India	0.48	0.40
Iran	0.42	0.44

1993–2002: average RBI. World RBI = 1

The number of citations, c , to papers in a given discipline for a given year increases with time elapsed since publication, t , approaching an asymptote c^* and roughly following an exponential function, $(1 - c/c^*) = \exp(-t/t_{1/2})$, where $t_{1/2}$ is a constant, the half-life of this function, which is reached when the citations have reached $c^*(1 - 1/e)$. The half-life is shorter in the biological sciences than in the physical sciences. An analysis² of papers published in 30 journals in the geological sciences between 1981 and 2002 yields a citation half-life of five years.

prints for the G8 countries excluding the United States are shown in Figure 3. The US citation share is about treble that of the United Kingdom, and is therefore compared separately on a different scale (Fig. 4), with the sum of the disciplinary footprints of the EU15 and the United Kingdom.

Figure 3 reveals some marked asymmetries. Russia is relatively strong in the physical sciences and engineering, and weak in the life sciences; Japan shows strengths in the physical sciences and engineering; France is strong in mathematics; Germany has the highest impact in the physical sciences; and the United Kingdom has the highest impact in this group in the medical, life and environmental sciences and is highly placed in mathematics, but does not show as strongly in the physical sciences and engineering. The comparison of the United States and the EU15 (Fig. 4) shows that the United States still has a bigger disciplinary footprint than the EU15, largely owing to its strength in the life sciences. The EU15 footprint is more symmetrical, being a little stronger than the United States in the physical sciences and engineering, but weaker in life and medical sciences.

Bang for the buck

May's 1997 analysis¹ included an evaluation of scientific output relative to government money spent on research and development (R&D), excluding defence; this analysis particularly gave rise to comment and discussion^{4–6}.

The following categories are the key factors in this analysis: higher education R&D, the output of which is people with research degrees and publications; government R&D, which generally gets more government money than higher education; publicly funded R&D, the sum of government and higher education; business funding of higher education R&D; and the gross domestic expenditure on R&D.

It is important to recognize that there are lags between changes in research funding and outputs (publications), and between outputs and their impact, and that multivariate models may be necessary to understand productivity. And there are international spillovers: raising one country's expenditure may increase output in other countries as well, particularly where direct collaboration occurs.

Figure 5 gathers data for the G8 nations, excepting Russia, into superimposed footprints. Relative strengths and weaknesses appear as asymmetries. The French and German footprints are symmetric and close to the average for this group. Other footprints show marked asymmetry. There is a much closer bunching of data points on the input side. The United Kingdom comes fourth on public R&D funding, close to the United States, and at sixth position on the higher-education measure, but first on all three normalized measures of output and outcome. Conversely, Japan is third on the public-funding measure, and first on higher

education, but seventh on all output and outcome indicators.

Clearly these metrics do not distinguish between public sector and business outputs. This would require a measurement of wealth generation from these inputs, for example, and is discussed in the following section. Within the G7 group, relative business funding of higher education R&D is highest in Japan and the United States, and lowest in Italy. Noting this important omission on inputs and outcomes, the footprints nevertheless give a snapshot of value-for-money measured by the international impact of the research output of each nation. The United Kingdom does remarkably well on these measures: top of the premier league, followed by Canada and the United States. I will return later to why this might be so.

Figure 6 compares the footprint for the EU15, including the strongly performing group of small nations in the EU, with the US and UK footprints. Although the research output and outcome metrics used show a

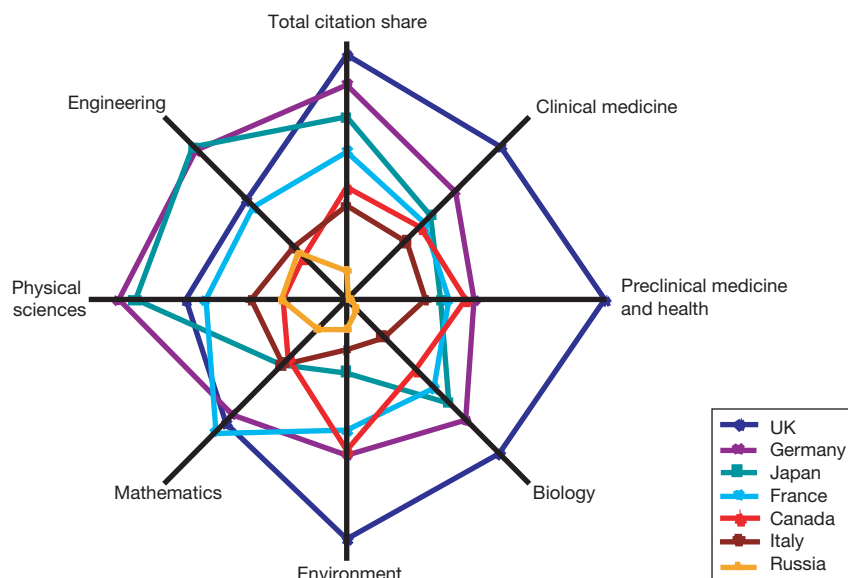


Figure 3 National strengths in different disciplines. Plot shows research footprints for the G8 nations excluding the United States, based on the national share of citations in each of seven disciplines and overall percentage share of citations. The distance from the origin to the data point is proportional to citation share. The medical and life sciences are shown to the right, mathematics and physical sciences to the left.

greater research impact in the EU, note that the citations per unit of publicly funded R&D are virtually identical.

Fuelling economic growth

It is also useful to consider indicators of business R&D activity. Table 3 shows the G8 nations ranked in order of business enterprise research and development expenditure as a percentage of GDP. This rank order differs in many respects from that in Table 1, which was based on research impact. In particular, Japan performs very strongly on this score, and the United States is significantly ahead of the EU15. The United Kingdom ranks a modest fifth on this important measure.

The knowledge base and economic fortunes of a nation are maintained and developed through the production of highly trained people. This is a necessary but not a sufficient condition for sustainable economic development: both political and macro-economic factors, including infrastructure investment, must also be in place. One measure of a nation's knowledge base is its output of PhD students (Table 3). Normalized to national populations, Germany, the United Kingdom and France are ahead of the United States on this measure, and Japan is relatively low down. Comparing these rankings with the figures for the proportion of the workforce in full-time employment as researchers reveals very different cultural practices (Table 3). On this measure the strong economic performances of Japan and the United States show clearly, taking first

and second places among the G8 nations for which data are available, well ahead of France, Germany, Canada and the United Kingdom. In Japan, industry tends to employ R&D personnel with first degrees and provide on-the-job training, with the expectation that employees will stay in the company all their working life. In Europe and the United States, in many industries, particularly outside engineering, it is more common to employ R&D personnel after PhD training. There is probably an advantage in the increased mobility of the R&D workforce with more generalized training up to PhD level, particularly during an economic downturn.

The private sector's annual investment in the public research sector is a useful measure of the interaction and knowledge transfer between business and higher education. This

indicator shows big changes over the ten-year period of this study. For the comparator group there was an average decline of 1.8% in business investment as a percentage of total public spending on R&D. The Italian investment as a percentage of publicly funded R&D fell from 3.4% to 0.6%; Poland also suffered a large fall. But Switzerland, the Netherlands and the United Kingdom showed large increases, the UK increase pulling it up from a low of 6.5% in 1995 to the current level of 11% of publicly funded R&D. This is the highest in the G8 group, followed by Canada and Germany at 7%, France at 5.5% and the United States at 3.5%.

An unequal world

The countries occupying the top eight places in the science citation rank order (Table 1) produced about 84.5% of the top 1% most cited publications between 1993 and 2001. The next nine countries produced 13%, and the final group share 2.5%. There is a stark disparity between the first and second divisions in the scientific impact of nations. Moreover, although my analysis includes only 31 of the world's 193 countries, these produce 97.5% of the world's most cited papers.

The political implications of this last comparison are difficult to exaggerate. South Africa, at 29th place in my rank ordering, is the only African country on the list. The Islamic countries are only represented by Iran at 30th, despite the high GDP of many of them and the prominence of some individuals, such as Nobel prizewinners Abdus Salam (physics, 1979) and Ahmed Zewail (chemistry, 1999)⁷.

My key point in response to these statistics is that sustainable economic development in highly competitive world markets requires a direct engagement in the generation of knowledge. Even modest improvements in healthcare, clean water, sanitation, food and transport need capabilities in engineering, technology, medicine, business, economics and social science beyond many countries' reach. The recent statements by

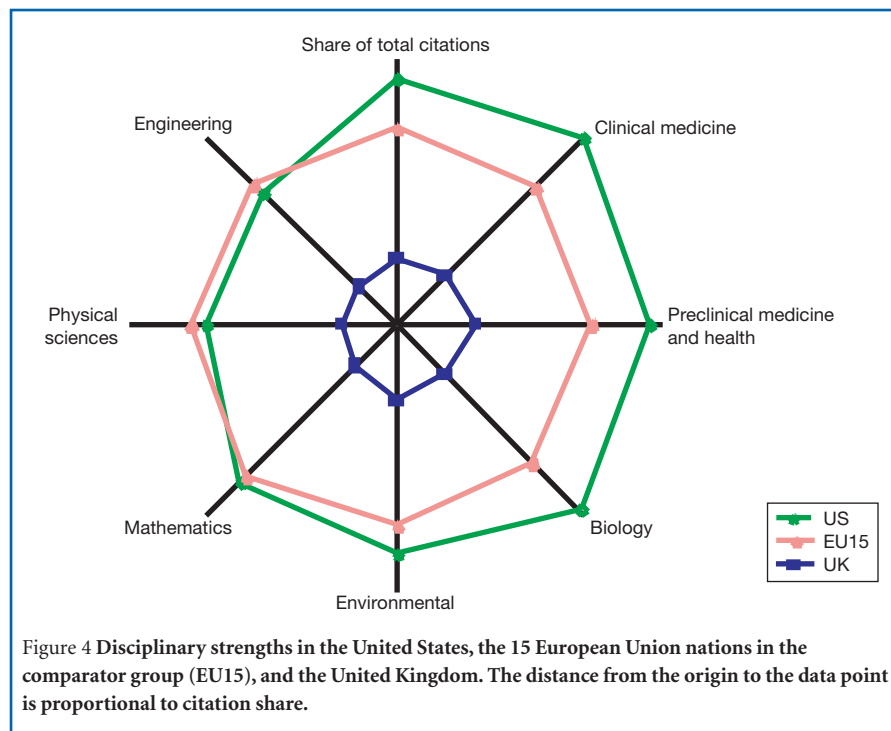
Table 3 Comparisons of private sector R&D spending and the output of PhDs and researchers*

Country	BERD†	BERD as % of GDP	PhDs	PhDs per head of population	Full-time researchers	Full-time researchers per 1,000 employed
Japan	65,726	2.12	10,962	0.08	644,208	9.59
US	169,228	1.97	44,955	0.17	1,148,271	8.17
Germany	31,013	1.66	24,940	0.30	238,944	5.93
France	18,186	1.38	10,056	0.17	156,004	5.99
UK	15,048	1.22	11,253	0.19	147,035	5.02
EU	95,733	1.19	6,323	0.18	784,066	5.6
Canada	8,343	1.06	3,871	0.13	90,245	5.88
Russia	6,577	0.72	-	-	-	-
Italy	6,569	0.53	3,494	0.06	69,621	3.09

All figures show average for 1997–2001, except PhDs, for which data are only available for 1998–2000.

*Researchers are defined as professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems and also in the management of the projects concerned, after the OECD/Frascati.

†Business enterprise research and development, in US\$ million at 1995 prices, adjusted for purchasing power.



of the United Kingdom (100), Germany (62), France (29), Switzerland (26), Sweden (17) and Italy (17). This result is at odds with research output and impact, but matches that based on highly cited papers. The considerably higher salaries that some US universities and institutions use to attract the world's top scientists is probably a major factor in sustaining this disparity. US universities' recent attempts to attract the young Russian mathematician Grigory Perelman illustrate this¹⁰. European nations should take note: particularly at the rarefied upper end, they are operating in a global market, and will need to improve salaries and conditions to compete for the top creative talent.

Top of the class

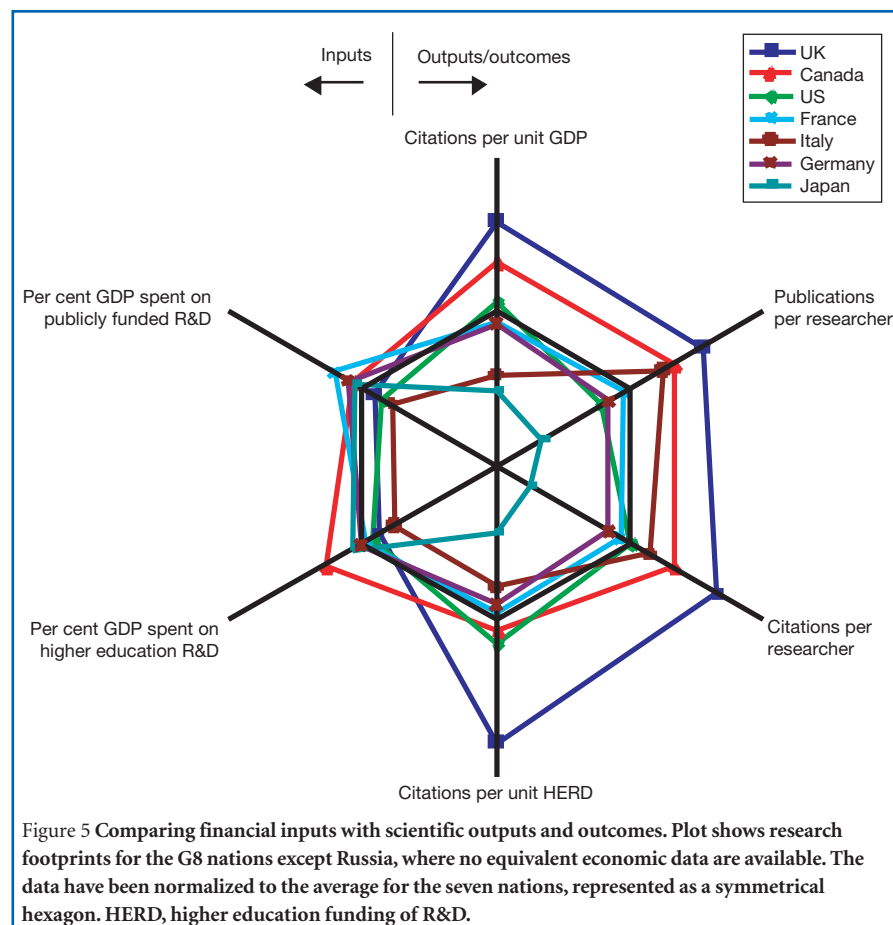
The Shanghai Institute of Education has recently published¹¹ a list of the top 500 world universities. The order is based on the number of Nobel laureates from 1911 to 2002, highly cited researchers, articles published in *Science* and *Nature*, the number of papers published and an average of these four criteria compared with the number of full-time faculty members in each institution. I believe none of these criteria are as reliable as citations. Based on an equal weighting for these five criteria, 15 (75%) of the top 20 universities are in the United States, four are in the United Kingdom, and

the UN secretary-general⁸ regarding the research and technology capacity of emergent nations amplify and emphasize these points. Nations exporting natural resources such as gold and oil can import technology and expertise, but only until these resources are exhausted. For them, sustainability should imply investment in alternative agricultural and technological capabilities through improvements in their skills base.

The cycles of poverty and dependence will only be broken by capacity-building between nations of high and low science intensity, often characterized as the North and the South. But it is important to note that simple citation rankings (Fig. 2, Tables 1 and 2) can hide important developments, particularly in countries such as China, placed 19th in my rank order, and India, at 22nd, which have developed their science base rapidly and effectively over the past few years. India's major science institutes have significant strength, produce high-quality graduates, and have made critical contributions to the country's sustained economic growth, as exemplified by the Bangalore software phenomenon. Similarly, Chinese universities have maintained high standards. With improved investment in research infrastructure and funding, China is attracting back scientists who have trained, and in some cases worked, in other countries, particularly the United States, and is sustaining the fastest economic growth in the world. National science outputs have not yet had a chance to catch up with these developments.

There have been several other recent comparisons of research activity. There is little consistency between broad measures of

research performance based on either the top four international science prizes¹ or the top 100 most cited individuals in each of 14 scientific fields⁹. Of the top 1,222 scientists from this last ranking, 815, or 66%, are from the United States and only 251 from the sum



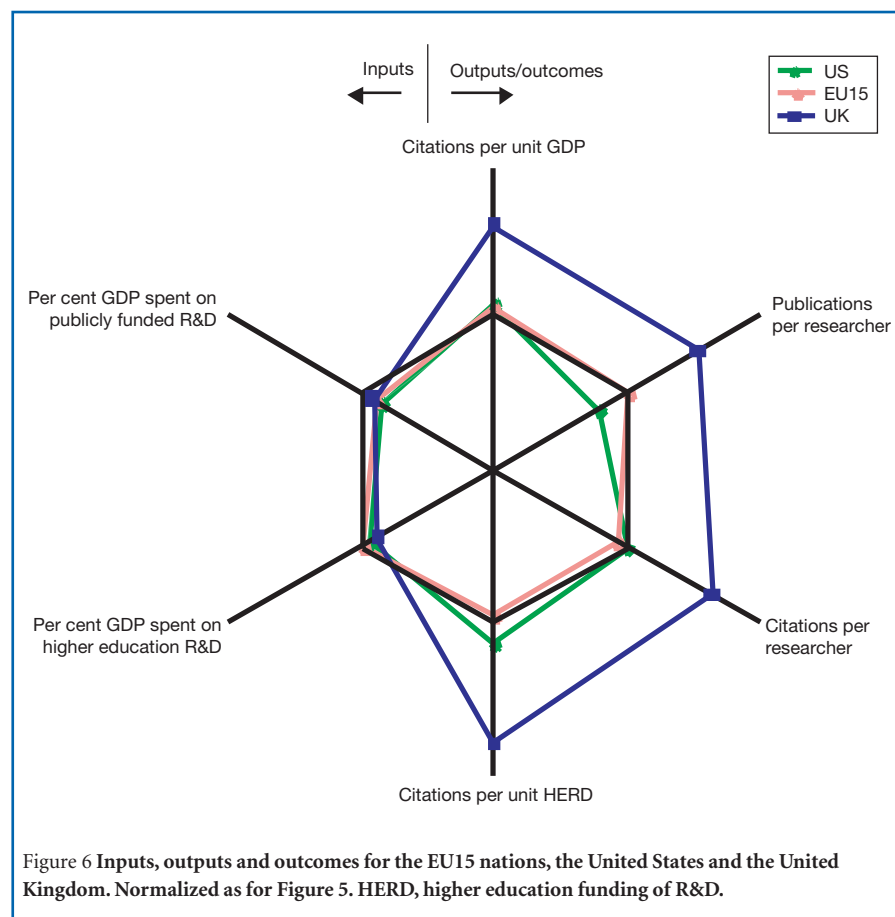


Figure 6 Inputs, outputs and outcomes for the EU15 nations, the United States and the United Kingdom. Normalized as for Figure 5. HERD, higher education funding of R&D.

one is in Japan. The United States also accounts for 75% of the top 40 institutions. In the top 100, there are 58 US and 31 European universities. The UK Research Assessment Exercise has produced a similar trend towards concentration of research excellence to that in the United States, with four institutions — Cambridge, Imperial College, Oxford and University College London — leading the UK research listings and stretching away from the others.

By contrast, comparing citations between the United States and the EU15 (Table 1 and Fig. 1) shows that the gap between the two has shrunk significantly since May's analysis based on figures up to 1993 (ref. 1). The EU now matches the United States in the physical sciences, engineering and mathematics, although it still lags in the life sciences (Fig. 4). The United Kingdom's share of science citations remains second only to the United States, and in intensity heads up the listing of large nations (Fig. 2). One paradoxical reason for the United Kingdom's current strength was the considerable cutback in public spending on science between 1980

and 1995. Although many UK scientists campaigned against these cuts, they encouraged a level of resourcefulness among researchers, and approaches to industry and the EU that are now bearing fruit. For instance, the United Kingdom's business investment in public research, as a proportion of public research R&D, is the highest in the world. And now that the present UK government is increasing funding and rebuilding infrastructure, the pruned plant of UK science is regrowing vigorously. The doubling of the Office of Science and Technology's science budget between 1997 and 2005 has been followed, as this article goes to press, with a treasury announcement that this budget will be further increased by an average annual growth of 5.6% per year, in real terms, over the three years to March 2008. This falls within a ten-year science and innovation strategy¹² for public and private UK R&D to reach 2.5% of GDP by 2014.

Among the other big EU players, Germany's contribution over the past 20 years has increased most sharply and is now within

a few per cent of the United Kingdom in its share of both citations and most-cited papers. Disciplinary footprints (Fig. 3) show a complementarity in science citations between Germany and the United Kingdom, with German strengths in physical sciences and engineering complementing UK strengths in medical, life and environmental sciences.

It is important also to recognize the role played in European science by the smaller northern countries, all of which rate highly in science intensity (Fig. 2). Thus, taking Belgium, Denmark, Finland, the Netherlands, Sweden and Switzerland together, with a total population of 53 million, in 1997–2001 this group generated 12.7% of the most cited papers, putting them in the same bracket as the United Kingdom (12.8%) and Germany (10.4%). Since the combined GDP of these countries is marginally (6%) smaller than the United Kingdom, their combined science citation intensity is higher.

A strong science base need not lead directly to wealth generation. For instance, although the strength of the UK science base has long been acknowledged, it has only recently begun to translate this into the development of high-tech clusters accompanying knowledge transfer between higher education and industry. However, strength in science has additional benefits for individual nations, and for the world as a whole. From global terrorism and the spread of disease to the dangers of global warming, we are increasingly facing the sorts of threats for which governments everywhere will need to turn to their scientists.

David A. King is at the Office of Science and Technology, 1 Victoria Street, London SW1H 0ET, UK.

1. May, R. M. *Science* **275**, 793–796 (1997).
2. *PSA target metrics for the UK Research Base* (UK Office of Science and Technology, October 2003); www.ost.gov.uk/policy/psa_target_metrics.htm. The original analysis provided by Evidence Ltd has been fully checked and updated for the work presented here. Publications and citation data are from Thomson Scientific.
3. Adams, J. *Nature* **396**, 615–618 (1998).
4. Grant, J. & Lewison, G. *Science* **278**, 878–879 (1997).
5. May, R. M. *Science* **278**, 879–880 (1998).
6. Margolis, R. M. & Kammen, D. M. *Science* **285**, 690–692 (1999).
7. Zewail, A. *Voyage Through Time* (American Univ. in Cairo Press, Cairo–New York, 2002).
8. Annan, K. *Science* **303**, 925 (2004).
9. Batty, M. *Env. Planning A* **35**, 761–765 (2003).
10. Singer, E. *Nature* **427**, 388–389 (2004).
11. <http://ed.sju.edu.cn/ranking.htm>
12. *Science and Innovation Investment Framework* (HM Treasury, London, 2004).

Acknowledgements

I am particularly grateful to Robert Thornes and John Kirk of the OST, London, for their comments and detailed checking of my figures, and to Zack King for his assistance in data mining.

Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue

Jesse M. Gray^{*1}, David S. Karow^{*2}, Hang Lu¹, Andy J. Chang¹, Jennifer S. Chang³, Ronald E. Ellis⁴ & Michael A. Marletta^{5,6} & Cornelia I. Bargmann¹

¹Howard Hughes Medical Institute and Departments of Anatomy and Biochemistry and Biophysics, The University of California, San Francisco, California 94143-0452, USA

²Graduate Program in Cellular and Molecular Biology, and ³Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

⁴UMDNJ School of Osteopathic Medicine, Stratford, New Jersey 08084, USA

⁵Departments of Chemistry and Molecular and Cell Biology, the University of California, Berkeley, CA 94720-1460, USA

⁶The Division of Physical Biosciences, Lawrence Berkeley National Lab, Berkeley, California 94720, USA

* These authors contributed equally to this work

Specialized oxygen-sensing cells in the nervous system generate rapid behavioural responses to oxygen. We show here that the nematode *Caenorhabditis elegans* exhibits a strong behavioural preference for 5–12% oxygen, avoiding higher and lower oxygen levels. 3',5'-cyclic guanosine monophosphate (cGMP) is a common second messenger in sensory transduction and is implicated in oxygen sensation. Avoidance of high oxygen levels by *C. elegans* requires the sensory cGMP-gated channel *tax-2/tax-4* and a specific soluble guanylate cyclase homologue, *gcy-35*. The GCY-35 haem domain binds molecular oxygen, unlike the haem domains of classical nitric-oxide-regulated guanylate cyclases. GCY-35 and TAX-4 mediate oxygen sensation in four sensory neurons that control a naturally polymorphic social feeding behaviour in *C. elegans*. Social feeding and related behaviours occur only when oxygen exceeds *C. elegans*' preferred level, and require *gcy-35* activity. Our results suggest that GCY-35 is regulated by molecular oxygen, and that social feeding can be a behavioural strategy for responding to hyperoxic environments.

All animals require oxygen as the essential electron acceptor in respiration, and respond to oxygen levels with behavioural and physiological changes. Soil, freshwater and marine animals encounter and avoid steep oxygen gradients in their natural environments^{1,2,3}. In mammals, oxygen acts through the hypoxia-inducible transcription factor HIF-1 to regulate erythropoietin production, red blood cell development, angiogenesis and cardiovascular physiology⁴. However, behavioural responses to oxygen occur much more rapidly than can be explained by changes in transcription. For example, the mammalian carotid body regulates ventilatory and circulatory responses to hypoxia within seconds⁵. The molecular nature of rapid oxygen sensation in the nervous system is not well understood.

GCY-35 mediates oxygen sensation

To examine oxygen-related behaviours in the nematode *C. elegans*, an aerotaxis assay was developed. Washed wild-type animals were placed in a gas-phase oxygen gradient from 0–21% that was produced by diffusion in a microdevice made of poly(dimethylsiloxane) (PDMS)⁶ and were allowed to move freely on an agar surface (Fig. 1a,b). Animals distributed themselves across the surface, avoiding low oxygen concentrations (<2%) as well as high oxygen concentrations (>12%) (Fig. 1c). The avoidance of hypoxia is consistent with previous studies⁷, but avoidance of hyperoxia has not previously been described, to our knowledge.

cGMP has been implicated in oxygen responses in *Drosophila*, where a nitric oxide (NO)-sensitive, cGMP-dependent kinase pathway mediates behavioural avoidance of hypoxia⁸. In cGMP second-messenger cascades, cGMP is produced from GTP by either membrane-bound guanylate cyclases or soluble guanylate cyclases (sGCs). All active sGCs characterized until now contain a haem cofactor and are activated by NO, which is produced by NO synthase. Notably, however, other haemoproteins can bind oxygen,

and prokaryotic haemoproteins mediate aerotaxis to preferred oxygen concentrations^{9,10}. The genome of the nematode *C. elegans* contains seven predicted sGC homologues (*gcy-31* through *gcy-37*) but no predicted NO synthase, suggesting that these cyclases might detect ligands other than endogenously-produced NO^{11,12}. The guanylate cyclases have the conserved histidine that ligates haem in mammalian β subunits (Supplementary Fig. 1)¹², and conservation of key catalytic residues from both α and β mammalian subunits suggests that *C. elegans* guanylate cyclases could be catalytically active (Supplementary Fig. 1)¹³. These genes were examined in more detail to determine whether they play a part in *C. elegans* oxygen sensing.

Previous studies demonstrated that *gcy-32* is expressed in URX, AQR and PQR sensory neurons and that *gcy-33* is expressed in BAG sensory neurons¹¹. The expression patterns of *gcy-34*, *gcy-35*, *gcy-36* and *gcy-37* were examined in transgenic animals bearing reporter genes in which upstream sequences for each gene were fused to sequences encoding green fluorescent protein (GFP). Each transgene was expressed in a small number of neurons (Fig. 2 and data not shown). Expression of *gcy-34*, *gcy-35*, *gcy-36* and *gcy-37* was consistently observed in URX, AQR, and PQR sensory neurons; *gcy-35* expression was also observed in ALN, SDQ and BDU neurons and variably in AVM, PLM and PLN neurons, pharyngeal and body wall muscles, and the excretory cell. The cells that reliably express *gcy-31*–*gcy-37* have the morphology of sensory neurons, suggesting a sensory role for the guanylate cyclase proteins, but the sensory cues that activate these cells are unknown.

To ask whether *gcy-35* might participate in oxygen sensing, a *gcy-35(ok769)* mutant from the *C. elegans* knockout consortium was characterized in the aerotaxis assay. *gcy-35(ok769)* deletes sequences corresponding to amino acids 456–545 of GCY-35, including key residues in the GC catalytic domain (Supplementary Fig. 1), and should abolish any ability of the *gcy-35(ok769)* gene product to

produce cGMP. In contrast to wild-type animals, *gcy-35(ok769)* animals avoided hypoxia but not hyperoxia in gas-phase oxygen gradients (Fig. 1d, Supplementary Fig. 2). This defect did not appear to be due to a general locomotory deficit, because *gcy-35* animals were mobile and chemotaxis-proficient (data not shown). The aerotaxis defect was rescued by expression of a *gcy-35* complementary DNA in URX, AQR and PQR (Fig. 1d, Supplementary Figs 2, 3). Thus *gcy-35* can act in URX, AQR and PQR sensory neurons to mediate avoidance of hyperoxic conditions.

cGMP can depolarize neurons by activating cyclic nucleotide-gated channels¹⁴. A cGMP-gated sensory transduction channel in *C. elegans* is composed of two subunits encoded by the *tax-2* and *tax-4* genes, which are co-expressed with *gcy-35* in URX, AQR and PQR neurons^{15,16}. To ask whether GCY-35 might act upstream of the cyclic nucleotide-gated channel, we tested *tax-4* and *tax-2* mutant and *tax-2; tax-4* double-mutant strains for aerotaxis behaviours. Like the *gcy-35* strain, *tax-4* and *tax-2* mutants failed to avoid hyperoxic conditions (Fig. 1e; Supplementary Figs 2, 3). The *tax-4* mutant defect was rescued by expression of *tax-4* in AQR, PQR and URX (Supplementary Figs 2, 3). These results indicate that the cGMP-gated channel is required for avoidance of hyperoxia, perhaps as the target for cGMP produced by GCY-35. In support of this model, a *tax-4; gcy-35* double mutant exhibited an aerotaxis defect resembling that of single mutants (Supplementary Figs 2, 3).

GCY-35 haem domain binds molecular oxygen

Gaseous ligands bind to sGCs through associated haem groups. To

investigate the potential ligand-binding characteristics of GCY-35, we cloned, expressed and purified the amino-terminal-predicted haem-binding fragment GCY-35(1–252). This protein was soluble (unlike full-length GCY-35) and tractable for biochemical analysis. Previous studies have shown that N-terminal haem-binding regions of the rat sGC $\beta 1$ subunit, $\beta 1(1-385)$ and $\beta 1(1-194)$, are spectroscopically similar to the full-length enzymes (data not shown and ref. 17). Therefore, the GCY-35(1–252) haem domain spectrum should be related to the ligand-binding characteristics of the full-length protein.

GCY-35(1–252) was characterized by ultraviolet/visible (UV/vis) spectroscopic analysis in the absence and presence of bound ligands. In an anaerobic environment, the purified protein was chemically treated with ferricyanide and dithionite to remove any ligands and to reduce the haem iron to its ferrous oxidation state. The Fe^{2+} -unligated, anaerobic spectrum of this protein exhibited a Soret maximum of 430 nm and a single, broad α/β region that was similar to ferrous-unligated sGC (Fig. 3a, c).

To test for oxygen binding, the unligated protein was exposed to air and immediately reanalysed by UV/vis spectroscopy. The resulting spectrum was characteristic of oxygen-bound haem, exhibiting a Soret maximum of 415 nm and a split α/β region similar to ferrous oxyhaemoglobin, indicative of a ferrous, low-spin complex (Fig. 3a, c). Like haemoglobin, GCY-35(1–252) was also able to bind NO and CO (Fig. 3c). In the presence of NO, GCY-35(1–252) exhibited a Soret maximum of 415 nm, similar to that of haemoglobin and a shoulder at 400 nm, similar to the Soret maximum of NO-bound

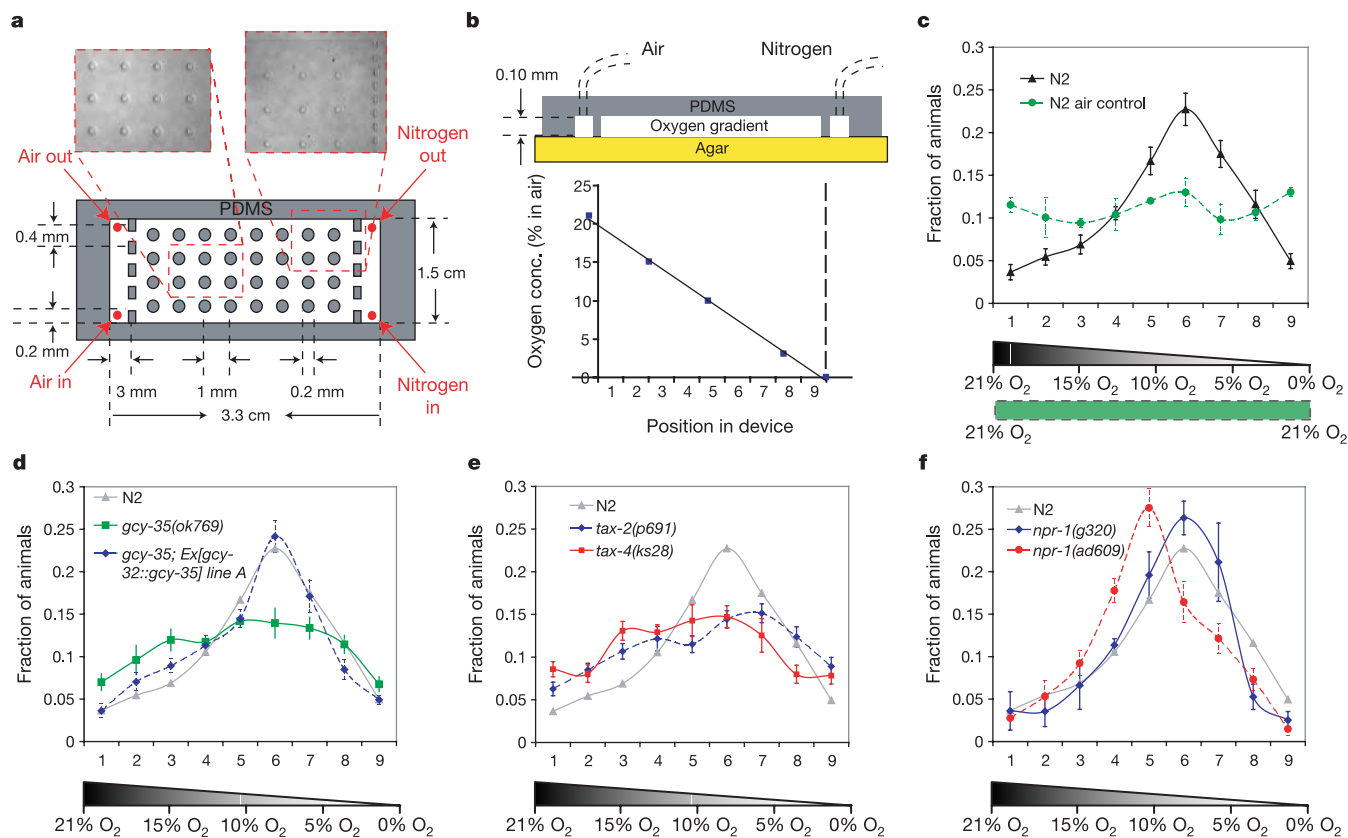


Figure 1 *gcy-35* mutants are defective in hyperoxia avoidance. **a**, Gas-phase PDMS aerotaxis device, top view. A gas-phase gradient was established by diffusion along the long axis of the device. **b**, Measured oxygen concentrations in aerotaxis device, and side view of device. **c**, Wild-type N2 animals accumulate at intermediate oxygen concentrations. **d**, Defective hyperoxia avoidance in *gcy-35* mutants, and rescue by

gcy-32::gcy-35 expression in URX, AQR, and PQR. **e**, Defective hyperoxia avoidance in *tax-4* and *tax-2* mutants. **f**, Similar oxygen preference of N2, *npr-1(g320)* and *npr-1(ad609)*. Error bars denote standard error of the mean (s.e.m.). For statistical analysis, see Supplementary Fig. 2.

sGC, suggesting that GCY-35(1–252) forms two stable nitrosyl complexes: a 5-coordinate high-spin complex that is similar to sGC and a 6-coordinate, low-spin complex that is similar to haemoglobin. The ligand-binding characteristics of the GCY-35 haem domain were most similar to oxygen-binding proteins like haemoglobin, and suggest that this protein could act as an oxygen sensor. No other native sGCs or sGC haem domain fragments characterized to date have been found to bind O₂. GCY-35(1–252) is unique in this respect.

Oxygen regulates social feeding behaviour

The URX, AQR and PQR neurons that co-express *gcy-35* and *tax-4* have previously been implicated in cGMP-mediated behaviours. The cyclic nucleotide-gated channel TAX-4 is required in these neurons to promote social feeding (or aggregation on a bacterial lawn) and bordering (the accumulation of animals on the thickest part of a bacterial lawn)¹⁸. Aggregation, bordering, burrowing into agar, and hyperactive locomotion represent a cluster of related behaviours that are not pronounced in the standard *C. elegans* laboratory strain, N2. However, social feeding behaviours are prominent in naturally isolated *C. elegans* strains that differ from N2 at the *npr-1* locus and in N2 strains that are deficient for the function of *npr-1*¹⁹. *npr-1* encodes a G-protein-coupled receptor for FMRF amide-like neuropeptides, and high levels of *npr-1* activity suppress aggregation and bordering behaviours²⁰. *tax-4* cGMP signalling stimulates aggregation and bordering by activating the URX, AQR and PQR neurons (as well as other neurons), whereas *npr-1* functions in URX, AQR and PQR to inhibit aggregation and bordering¹⁸. These reciprocal results suggest that the activity of URX, AQR and PQR regulates aggregation and bordering behaviours.

If the guanylate cyclases expressed in URX, AQR and PQR are molecular oxygen sensors, then oxygen should regulate the activity of these cells. To test this hypothesis, aggregation and bordering behaviours were examined in animals exposed to a constant flow of gas with different concentrations of oxygen. Initial experiments

were conducted by shifting animals from 21% to 7% oxygen, the concentration that was preferred by *C. elegans* in aerotaxis experiments. *npr-1(ad609)* is a loss-of-function mutation, and *npr-1(g320)* is the reduced-function allele of *npr-1* present in natural social strains¹⁹. *npr-1(ad609)* mutants shifted to 7% oxygen rapidly suppressed both aggregation and bordering behaviours (Fig. 4b, d, f). Suppression was evident within three minutes of shifting to 7% oxygen and was stable for at least thirty minutes after the shift (Fig. 4d, f). A return to 21% oxygen led to the reappearance of aggregation and bordering behaviours within three minutes (Fig. 4d, f). Similar effects were seen with *npr-1(g320)* (data not shown).

A smaller shift from 21% oxygen to 15% oxygen or 10% oxygen led to a similar, but less marked suppression of aggregation and bordering behaviours (Fig. 5a, b). In all cases, the change in behaviour was reversed by returning to 21% oxygen. Thus decreases in oxygen lead to a dose-dependent suppression of social feeding behaviour, suggesting that oxygen serves as a quantitative regulator of social feeding by URX, AQR and PQR.

Bacteria alter oxygen levels and responses

All behaviours in the social feeding cluster are most pronounced in the presence of bacterial food¹⁹. The aerotaxis assay is conducted in the absence of food, and under these circumstances *npr-1* strains exhibited aerotaxis with preferred concentrations similar to those preferred by N2 animals (Fig. 1f). Because social behaviours are food-induced, we also tested aerotaxis in N2 and *npr-1* strains in the presence of food (Fig. 5c, Supplementary Fig. 2). On a thin bacterial lawn, aerotaxis behaviours in N2 were blunted, with a greatly

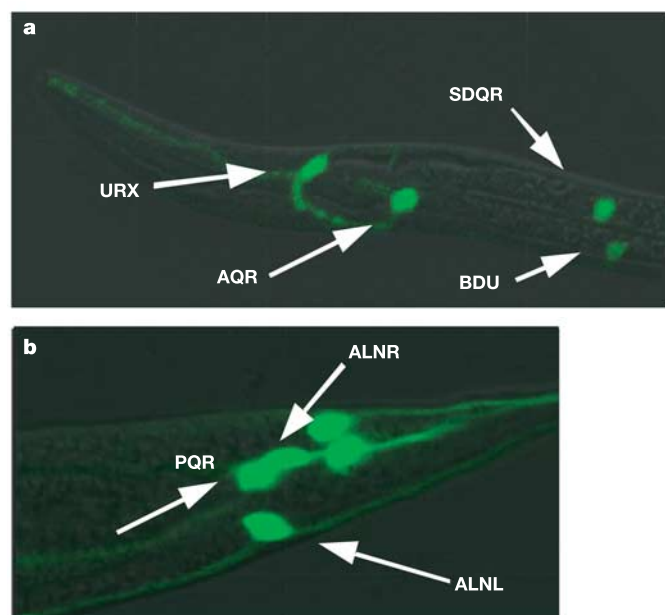


Figure 2 *gcy-35::gfp* is expressed in URX, AQR, PQR and other sensory neurons. **a**, Lateral view of the anterior body showing URX, AQR, SDQR and BDU neurons. Anterior is at left and ventral is down. **b**, Ventral view of the tail showing PQR and ALNL/R neurons. The more posterior cells may be PLM neurons, the sisters of ALNL or ALNR, or PLN neurons.

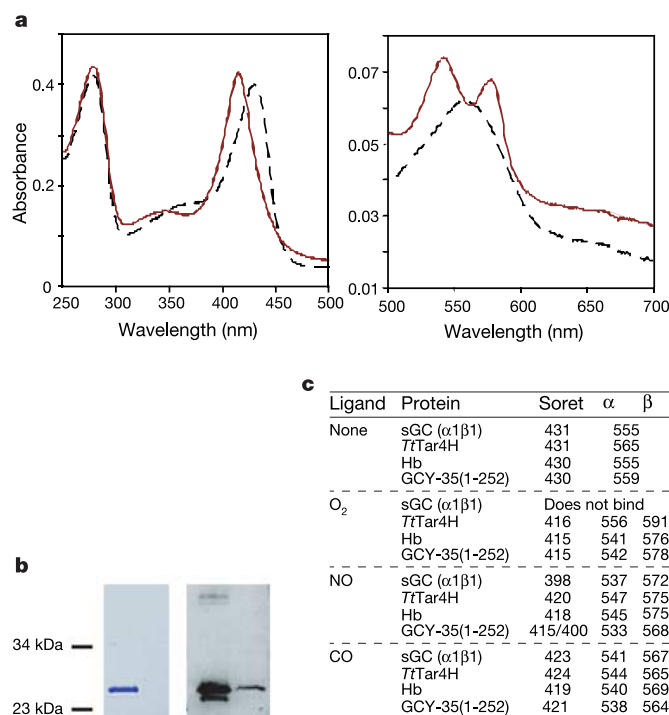


Figure 3 Characterization of GCY-35(1–252) binding to gases. **a**, UV/vis spectroscopy of GCY-35(1–252). The black broken trace shows anaerobic spectrum of ferrous-unligated complex (peaks at 430 and 559 nm). The solid red trace shows the same sample after exposure to air, and is indicative of a ferrous-oxy complex (peaks at 415, 542 and 578 nm). **b**, Purification of GCY-35(1–252):His. Left, Coomassie blue-stained gel. Right, western blot with affinity-purified anti-GCY-35 antisera, lanes at tenfold different dilutions. **c**, Comparison of spectroscopic data for GCY-35(1–252), soluble guanylate cyclase³², haemoglobin³³, and *Thermoanaerobacter tengcongensis* Tar4H²⁵ unligated and in the presence of CO, NO and O₂.

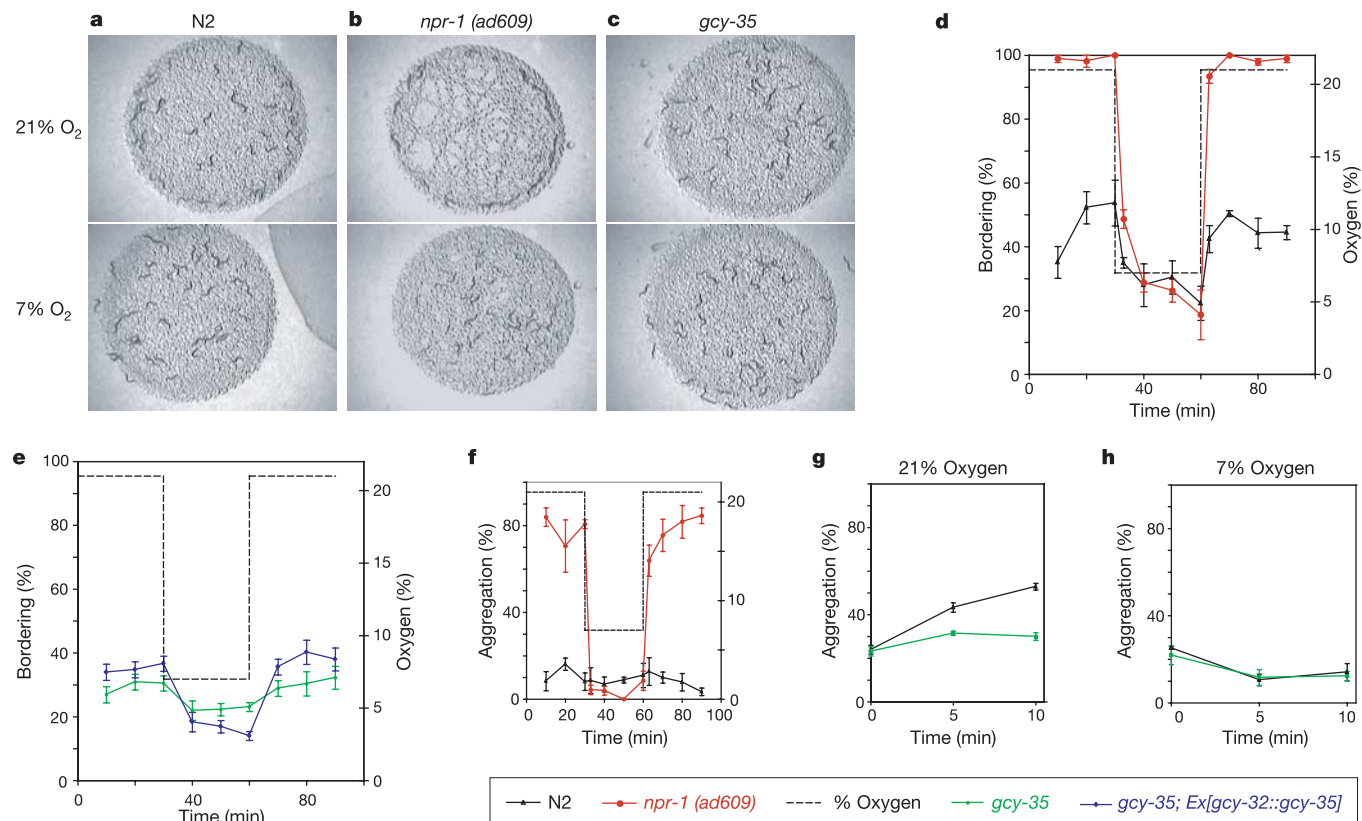


Figure 4 Oxygen stimulates GCY-35-dependent aggregation and bordering. **a–c**, Animals equilibrated at 21% oxygen or 7% oxygen for 30 min. **a**, N2; **b**, *npr-1*(ad609); **c**, *gcy-35*(ok769). **d–f**, Oxygen shifts from 21%–7%–21% (dotted lines). **d**, **f**, 7% oxygen suppresses bordering in N2 and *npr-1* strains and aggregation in *npr-1*

strains. **e**, Reduced oxygen sensitivity of bordering in *gcy-35*(ok769), and rescue by *gcy-32::gcy-35* transgene. **g**, **h**, Aggregation of N2 and *gcy-35*(ok769) immediately after transfer to a fresh bacterial lawn in 21% or 7% oxygen. Error bars denote s.e.m. For statistical analysis see Supplementary Information.

reduced avoidance of hyperoxia. By contrast, *npr-1* strains exhibited robust hyperoxia avoidance in the presence or absence of food.

How does oxygen sensation relate to the cluster of social feeding behaviours? We observed that the thick lawns of *E. coli* that are usually fed to *C. elegans* consume oxygen more quickly than oxygen diffuses through the lawn. The border of a thick lawn has an effective oxygen concentration of 12.8%, compared to 17.1% in the centre of the lawn (Fig. 5d). Thus bordering behaviours may be caused, in part, by the strong preference of *npr-1* strains for lower oxygen concentrations.

Moderate bordering behaviour is evident in N2 animals grown on thick bacterial lawns, consistent with the moderate hyperoxia avoidance that N2 exhibits on food. N2 bordering behaviour was suppressed by a shift from 21% to 7% oxygen (Fig. 4a, d) or from 21% to other low oxygen concentrations (Fig. 5a). Conversely, bordering in the N2 background was stimulated by a shift from 7% to 21% oxygen (Fig. 4d). N2 animals aggregate for the first 20 min after transfer to a fresh bacterial lawn (that is, a lawn with no worms on it)²¹. This aggregation was suppressed at 7% oxygen (Fig. 4g, h). Thus oxygen acts in parallel to *npr-1*, regulating social feeding behaviours regardless of the *npr-1* genotype of the animal.

In the N2 genetic background, *gcy-35* mutants exhibited lower levels of aggregation and bordering than wild-type animals (Fig. 4c, e, g). Changes in oxygen concentration between 21% and 7% had little effect on bordering behaviour in *gcy-35* mutants (Fig. 4e). Moreover, *gcy-35* mutants aggregated less than N2 when placed on a fresh bacterial lawn under 21% oxygen (Fig. 4g). Expression of a *gcy-35* cDNA in URX, AQR and PQR restored the oxygen sensitivity of these mutants (Fig. 4e). These results suggest that oxygen acts

through *gcy-35* to regulate aggregation and bordering, a process that is antagonized by *npr-1* activity. Indeed, an independent study recently showed that *gcy-35; npr-1* double mutants do not border or aggregate²². Like *gcy-35* mutants, *tax-4* mutants responded poorly to changes in oxygen levels, suggesting that oxygen regulation of social feeding behaviour depends on cGMP-gated channels (Supplementary Fig. 3).

Discussion

These behavioural results demonstrate that GCY-35 is required for avoidance of hyperoxia and for oxygen-induced aggregation and bordering. Biochemical evidence suggests that GCY-35 forms a stable ferrous-oxy complex. Canonical NO-sensitive sGCs do not bind oxygen; indeed, their inability to bind oxygen is essential to their ability to sense NO, given that the NO:O₂ ratio in tissues is about 1:1,000 (refs 23, 24). By analogy with the NO-sensitive sGCs, we suggest that oxygen modulates GCY-35 guanylate cyclase activity, either directly or by competing with other activators like NO. Although catalytic activity remains to be demonstrated, several bacterial and archaeal haem domains mediate aerotaxis⁹, and one other sGC-like haem domain, Tar4H from *Thermoanaerobacter tengcongensis*, has been found to bind oxygen²⁵. Our results suggest that the soluble guanylate cyclase GCY-35 may represent one member of a new class of oxygen-sensitive sGCs.

GCY-35 mediates oxygen sensing in URX, AQR and PQR. The URX sensory neurons have dendrites that extend to the tip of the nose, suggesting that URX detects external stimuli. The AQR and PQR neurons extend dendrites with ciliated endings into the pseudocoelom, an internal, fluid-filled cavity immediately under

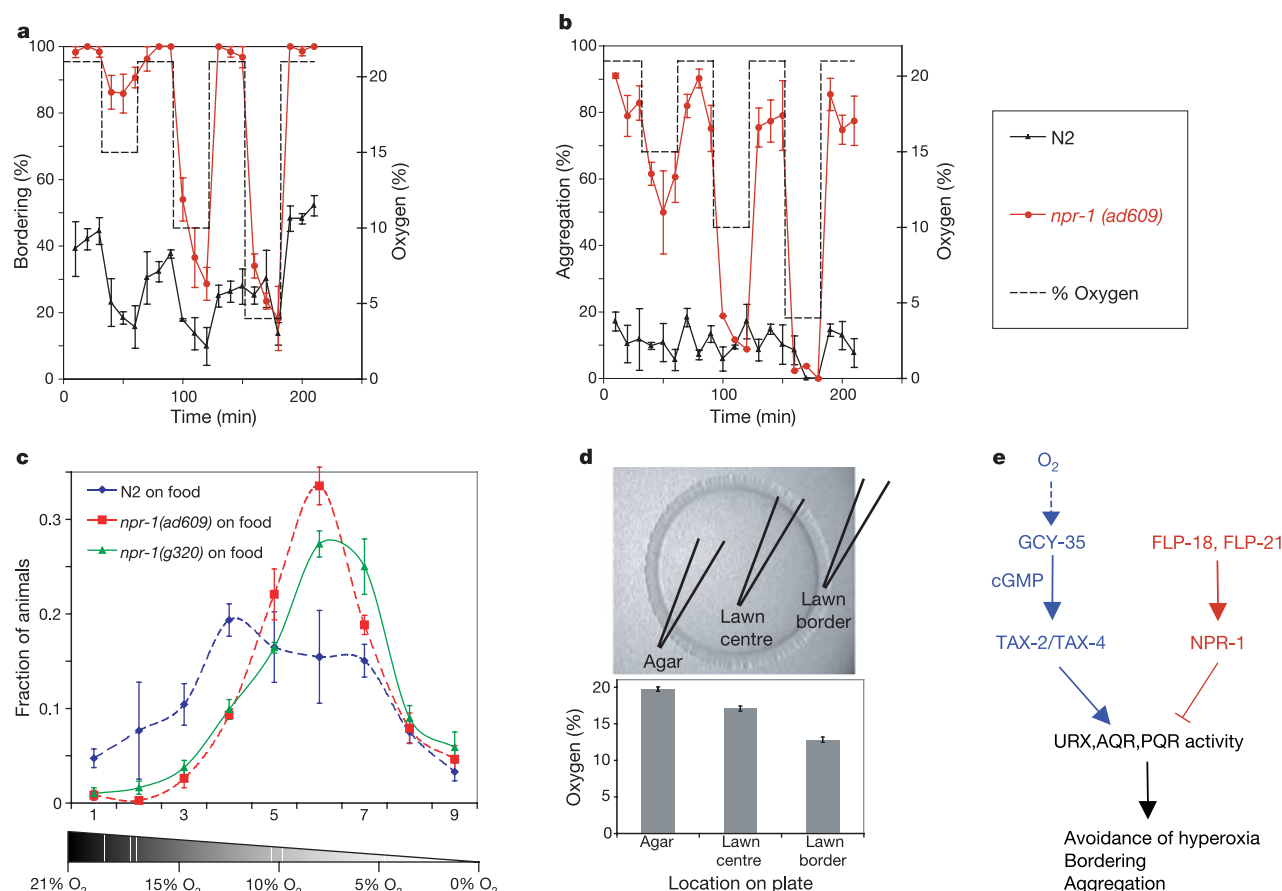


Figure 5 Regulation of oxygen responses by food. **a, b**, Bordering and aggregation during oxygen shifts (dotted line). **c**, Aerotaxis on a thin bacterial lawn. N2 aerotaxis is suppressed. **d**, Oxygen concentrations in and near a thick OP50 bacterial lawn. Error bars denote s.e.m. **e**, Model for oxygen regulation of behaviour. Oxygen directly or indirectly

regulates GCY-35; cGMP activates the TAX-2/TAX-4 cGMP-gated channel in URX, AQR and PQR to promote hyperoxia avoidance, bordering and aggregation. The neuropeptides FLP-18/21 (possibly released by pharyngeal neurons during feeding²⁰) activate the neuropeptide receptor NPR-1, which inhibits URX, AQR and PQR.

the epithelium²⁶. AQR and PQR may detect internally generated stimuli or internal levels of externally produced stimuli. On the basis of its rate of diffusion through water, molecular oxygen should diffuse from the environment to the AQR and PQR sensory endings in less than a second.

Because oxygen diffuses rapidly through small animals like *C. elegans*, it does not become a limiting factor for respiration until it reaches external concentrations below 4% (refs 27, 28). Above 4% oxygen, additional increases in oxygen do not affect respiration, but are likely to increase cellular oxidative damage, because the propensity to produce reactive oxygen species will increase with oxygen concentration²⁹. One mechanism for protecting against such damage may be a behavioural preference for lower oxygen environments. This strategy has been well characterized in bacteria, which distribute themselves along oxygen gradients in characteristic bands at their preferred concentration³⁰. In *C. elegans*, behavioural preference for lower oxygen concentrations could also result from factors unrelated to oxidative stress; for example, low oxygen may signal the presence of food in the form of actively growing bacteria that consume oxygen more quickly than the oxygen can diffuse through soil.

Our results suggest that social feeding behaviour represents an integrated behavioural response to aversive hyperoxic conditions (Fig. 5e). Accumulation at the border of the lawn may represent direct aerotactic avoidance of hyperoxia. By contrast, small groups of animals are unlikely to create oxygen gradients sufficient to

attract more animals. We propose that high oxygen levels and other aversive chemical stimuli²¹ serve as sensory triggers that can initiate social behaviour by activating chemotaxis or mechanotaxis to other animals.

An independent study found that mutations in either *gcy-35* or the related gene *gcy-36* suppress bordering and aggregation in *npr-1* mutants²². Together, these results suggest that oxygen regulates social behaviour by modulating the guanylate cyclase activity of GCY-35 and GCY-36. It will be interesting to see whether soluble guanylate cyclases act as neuronal oxygen sensors in other animals. □

Methods

Standard techniques used for genetics, molecular biology, biochemistry and statistics are described in the Supplementary Methods.

Oxygen binding of bacterially-produced GCY-35(1–252)

Purified bacterially produced protein was made anaerobic in an O_2 -scavenged gas train with ten cycles of alternate evacuation and purging with purified argon and brought into an anaerobic glove bag. Ferricyanide (~100 equivalents) was added and then removed using a PD10 desalting column that had been equilibrated with 50 mM TEA pH 7.5 and 50 mM NaCl (Buffer C). The protein was then reduced using dithionite (~100 equivalents). The dithionite was then removed in the same manner. A ferrous-unligated UV/vis absorption spectrum was recorded in an anaerobic cuvette on a Cary 3E spectrophotometer equipped with a Neslab RTE-100 temperature controller set at 10 °C. Spectra were recorded from protein in Buffer C. Fe^{2+} - O_2 protein and other gas-bound proteins were generated by exposing Fe^{2+} -unligated GCY-35(1–252) to air or other gases before recording a spectrum. Bacterial haem can assemble with recombinant eukaryotic haem-binding proteins, and because *C. elegans* cannot itself synthesize haem³¹, bacterial

haem is probably the natural form for *C. elegans* guanylate cyclases.

Behavioural assays

For gas-phase aerotaxis assays, microdevices were fabricated using the PDMS rapid prototyping technique⁶. The photolithography masks were laser-printed on silver halide films with 0.625-µm resolution, and used to produce the prototype masters in a photo-patternable epoxy (SU-8-50, Microchem) on silicon wafers using standard UV photolithography. The masters were silanized using vapour-phase tridecafluoro-1,1,2,2-tetrahydrooctyl trichlorosilane (United Chemical Technologies). The PDMS devices were micromoulded using two-part Sylgard 184 silicone elastomer (Dow Corning) against the masters. Before the assay, the devices were cleaned in ethanol followed by deionized water and dried overnight at 65 °C.

In each assay, 30–200 washed adult animals were placed on an agar plate before the device was placed over them. Air and nitrogen gas (at ~21–22 °C, 1 atm) were delivered to the source and drain chambers under laminar flow at 1 ml min⁻¹ for 15–30 min from gas-tight syringes using a syringe pump (PHD2000, Harvard Apparatus). Although PDMS and agar are both oxygen-permeable, the diffusion rate of gases through these media is substantially smaller than the gas flux between the source and drain, and did not disrupt the oxygen gradient in gas phase. Animals at nine equally sized regions in the device were counted at the end of the assay, determined by at least two consecutive scorings (five minutes apart), yielding similar spatial distributions. Each aerotaxis data point represents 3–8 assays with 80 or more animals per assay. In assays with fewer than 80 total animals, counts from experiments done on the same day were combined for data analysis. The oxygen concentration gradient in the agar was measured using a Clark-style oxygen microelectrode with guard cathode (Chemical microsensor #1201, Diamond Microsensors), which was calibrated with standards immediately before each use.

For aggregation and bordering assays, 40 animals were picked onto a bacterial lawn, allowed to equilibrate for an hour, placed into a flow chamber with a constant gas flow of 150 ml min⁻¹, and exposed to 21% oxygen for at least 30 min. Oxygen was then adjusted to different concentrations for various periods of time at a constant flow rate of 150 ml min⁻¹. The flow chamber was a 100 mm × 15 mm Petri dish with a modified cover: two female luers (Biorad) were melted and glued into the ends of the cover. One luer served as the gas inlet and was connected to tubing, while the other luer was always exposed to air and served as the outlet. Gas mixtures with varying concentrations of oxygen were obtained by mixing oxygen and nitrogen (99.997%) in a flow meter (Cole-Parmer). All gases were obtained from Airgas. Images were captured every 10 min with an Ultrachip CCTV camera (JE-7442, Javelin) mounted on a stereomicroscope (Wild M3Z, Leica). The analogue camera output was connected to the RCA port on a Macintosh Power PC 7600/132 equipped with an on-board analogue/digital converter. Movies and photographs were captured with Adobe Premiere software (version 4.0.1). Bacterial lawns (OP50 strain) were grown for four days before the assay and were 10–13 mm in diameter. Animals within 1 mm of the edge of the bacterial lawn were scored to be in the border. Any animal that was touching at least one other animal across at least 30% of its body was scored to be in an aggregate.

For the transient aggregation experiments in Fig. 4g, h, animals were placed in the flow chamber and counted immediately after transfer to a fresh bacterial lawn. Fresh bacterial lawns were 4-day-old strain OP50 lawns that had never encountered worms.

In Fig. 5c, a thin bacterial lawn (bacterial strain OP50) was produced by seeding 10-cm NGM plates that were used 8–10 h later.

Received 1 March; accepted 7 June 2004; doi:10.1038/nature02714.

Published online 27 June 2004.

1. Wu, R. S. Hypoxia: from molecular responses to ecosystem responses. *Mar. Pollut. Bull.* **45**, 35–45 (2002).
2. Wannamaker, C. M. & Rice, J. A. Effects of hypoxia on movements and behavior of selected estuarine organisms from the southeastern United States. *J. Exp. Mar. Biol. Ecol.* **249**, 145–163 (2000).
3. Sylvia, D. M., Fuhrmann, J. J., Hartel, P. G. & Zuberer, D. A. *Principles and Applications of Soil Microbiology* (Prentice Hall, Upper Saddle River, New Jersey, 1998).
4. Semenza, G. L. HIF-1: mediator of physiological and pathophysiological responses to hypoxia. *J. Appl. Physiol.* **88**, 1474–1480 (2000).
5. Lopez-Barneo, J. Oxygen and glucose sensing by carotid body glomus cells. *Curr. Opin. Neurobiol.* **13**, 493–499 (2003).
6. Duffy, D. C., McDonald, J. C., Schueller, O. J. A. & Whitesides, G. M. Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). *Anal. Chem.* **70**, 4974–4984 (1998).
7. Dusenbery, D. B. Appetitive response of the nematode *Caenorhabditis elegans* to oxygen. *J. Comp. Physiol.* **136**, 333–336 (1980).
8. Wingrove, J. A. & O'Farrell, P. H. Nitric oxide contributes to behavioral, cellular, and developmental responses to low oxygen in *Drosophila*. *Cell* **98**, 105–114 (1999).
9. Hou, S. *et al.* Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature* **403**, 540–544 (2000).

10. Jain, R. & Chan, M. K. Mechanisms of ligand discrimination by heme proteins. *J. Biol. Inorg. Chem.* **8**, 1–11 (2003).
11. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
12. Morton, D. B., Hudson, M. L., Waters, E. & O'Shea, M. Soluble guanylyl cyclases in *Caenorhabditis elegans*: NO is not the answer. *Curr. Biol.* **9**, R546–R547 (1999).
13. Morton, D. B. Invertebrates yield a plethora of atypical guanylyl cyclases. *Mol. Neurobiol.* **29**, 97–116 (2004).
14. Finn, J. T., Grunwald, M. E. & Yau, K. W. Cyclic nucleotide-gated ion channels: an extended family with diverse functions. *Annu. Rev. Physiol.* **58**, 395–426 (1996).
15. Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans*. *Neuron* **17**, 695–706 (1996).
16. Komatsu, H., Mori, I., Rhee, J. S., Aiakie, N. & Ohshima, Y. Mutations in a cyclic nucleotide-gated channel lead to abnormal thermosensation and chemosensation in *C. elegans*. *Neuron* **17**, 707–718 (1996).
17. Zhao, Y. & Marletta, M. A. Localization of the heme binding region in soluble guanylate cyclase. *Biochemistry* **36**, 15959–15964 (1997).
18. Coates, J. C. & de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*. *Nature* **419**, 925–929 (2002).
19. de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).
20. Rogers, C. *et al.* Inhibition of *Caenorhabditis elegans* social feeding by FMRFamide-related peptide activation of NPR-1. *Nature Neurosci.* **6**, 1178–1185 (2003).
21. de Bono, M., Tobin, D. M., Davis, M. W., Avery, L. & Bargmann, C. I. Social feeding in *Caenorhabditis elegans* is induced by neurons that detect aversive stimuli. *Nature* **419**, 899–903 (2002).
22. Cheung, B. H. H., Arellano-Carbajal, E., Rybicki, I. & de Bono, M. Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. elegans* aggregation behaviour. *Curr. Biol.* (in press).
23. Malinski, T. & Taha, Z. Nitric oxide release from a single cell measured in situ by a porphyrinic-based microsensor. *Nature* **358**, 676–678 (1992).
24. Malinski, T., Taha, Z. & Grunfeld, S. Diffusion of nitric oxide in the aorta wall monitored in situ by porphyrinic microsensors. *Biochem. Biophys. Res. Commun.* **193**, 1076–1082 (1993).
25. Karow, D. S. *et al.* Spectroscopic characterization of the sGC-like heme domains from *Vibrio cholerae* and *Thermotoga maritima*. *Biochemistry* (in press).
26. White, J., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
27. Van Voorhies, W. A. & Ward, S. Broad oxygen tolerance in the nematode *Caenorhabditis elegans*. *J. Exp. Biol.* **203**, 2467–2478 (2000).
28. Anderson, G. L. & Dusenbery, D. B. Critical oxygen tension of *Caenorhabditis elegans*. *J. Nematol.* **9**, 253–254 (1977).
29. Imlay, J. A. Pathways of oxidative damage. *Annu. Rev. Microbiol.* **57**, 395–418 (2003).
30. Barak, R., Nur, I., Okon, Y. & Henis, Y. Aerotactic response of *Azospirillum brasilense*. *J. Bacteriol.* **152**, 643–649 (1982).
31. Hieb, W. F., Stokstad, E. L. & Rothstein, M. Heme requirement for reproduction of a free-living nematode. *Science* **168**, 143–144 (1970).
32. Stone, J. R. & Marletta, M. A. Soluble guanylate cyclase from bovine lung: activation with nitric oxide and carbon monoxide and spectral characterization of the ferrous and ferric states. *Biochemistry* **33**, 5636–5640 (1994).
33. Di Iorio, E. E. In *Hemoglobins* (eds Antonini, E., Rossi-Bernardi, L. & Chiancone, E.) 57–71 (Academic, New York, 1981).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Feldman for discussions and contributions to the aerotaxis assay, M. Hudson and D. Morton for advice and discussions, C. Ross, S. Nicholls and M. Miazgowiec for technical assistance, M. Zimmer for the *gcy-32* promoter, S. McCarroll for the pSM1 vector, the *C. elegans* Knockout Consortium and *Caenorhabditis* Genetics Center (CGC) for the *gcy-35(ok769)* mutant strain, and B. Cheung and M. de Bono for sharing their results before publication. J.M.G. was supported by a Howard Hughes Medical Institute Predoctoral Fellowship. A.J.C. was supported by an NSF Predoctoral Fellowship. C.I.B. is an Investigator of the Howard Hughes Medical Institute. This work was supported by funding from the Howard Hughes Medical Institute (to C.I.B.) and by the LDRD fund from the Lawrence Berkeley National Lab (to M.A.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.I.B. (cori@itsa.ucsf.edu) or M.A.M. (marletta@berkeley.edu).

Fast delivery of meteorites to Earth after a major asteroid collision

Philipp R. Heck¹, Birger Schmitz², Heinrich Baur¹, Alex N. Halliday¹ & Rainer Wieler¹

¹ETH Zürich, Isotope Geology, NO C61, CH-8092 Zürich, Switzerland

²University of Lund, Department of Geology, Sölvegatan 12, SE-22362 Lund, Sweden

Very large collisions in the asteroid belt could lead temporarily to a substantial increase in the rate of impacts of meteorites on Earth. Orbital simulations predict that fragments from such events may arrive considerably faster than the typical transit times of meteorites falling today, because in some large impacts part of the debris is transferred directly into a resonant orbit with Jupiter^{1,2}. Such an efficient meteorite delivery track, however, has not been verified. Here we report high-sensitivity measurements of noble gases produced by cosmic rays in chromite grains from a unique suite of fossil meteorites³ preserved in ~480 million year old sediments. The transfer times deduced from the noble gases are as short as ~10⁵ years, and they increase with stratigraphic height in agreement with the estimated duration of sedimentation. These data provide powerful evidence that this unusual

meteorite occurrence was the result of a long-lasting rain of meteorites following the destruction of an asteroid, and show that at least one strong resonance in the main asteroid belt can deliver material into the inner Solar System within the short timescales suggested by dynamical models.

One of the best-documented large asteroid collisions in the meteorite record disrupted a parent body of the L chondrites, one of the two most abundant meteorite classes. Many L chondrites show signs of severe shock and were degassed about 500 million years (Myr) ago, as demonstrated by their K-Ar gas retention ages^{4,5}. The L chondrites from this event that are falling today were delivered to Earth by later collisions mostly some 3–60 Myr ago, as measured by the amounts of noble gases produced by cosmic rays (cosmogenic noble gases). Evidence that this event also delivered many meteorite-sized fragments shortly after the original collision has been found in the form of an unexpectedly high abundance of fossil meteorites in ~480 Myr old sediments in southern Sweden³. The meteorite flux recorded in these sediments was two orders of magnitude higher than today⁶. Apart from tiny relict chromite grains, the minerals in the fossil meteorites are almost completely replaced by diagenetic pseudomorphs. The chemical composition of the chromite grains matches that of modern L (or LL) chondrites, providing strong evidence that the fossil meteorites originate from the same major collision as the one recorded in the low gas retention ages of L chondrites. This conclusion has been corroborated by the finding of extraterrestrial chromite grains of L (or LL) chondrite composition dispersed in ~480 Myr old sediments in other locations in southern Sweden⁶.

Extraterrestrial ³He is present in bulk samples of the fossil meteorites, but the very low concentrations indicate a nearly complete loss during diagenesis⁷. We therefore have measured He and Ne isotopes in chromite grains from nine fossil meteorites from six different sediment beds with the high sensitivity compressor source noble gas mass spectrometer at ETH⁸. Figure 1 shows the positions of the meteorites in the sediment column. The total deposition time from the lowermost Arkeologen (Ark) bed to the uppermost Glaskarten (Gla) bed is probably between 1 and 2 Myr (ref. 9), with the true value probably being closer to the low end of this interval.

He and Ne data are shown in Table 1 and Supplementary Table 1.

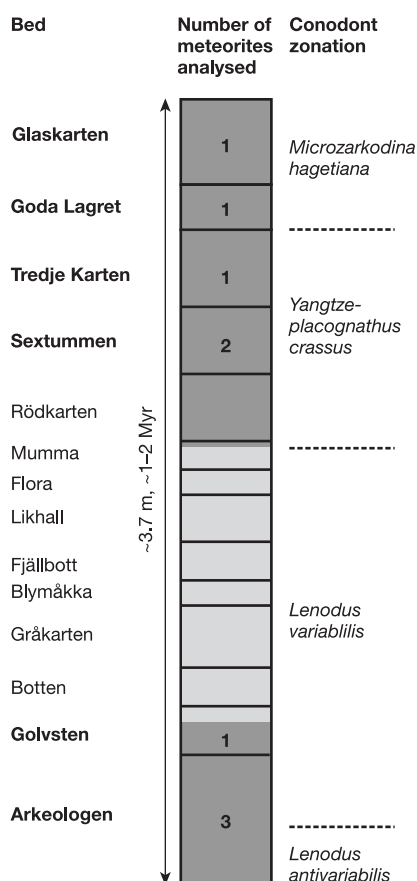


Figure 1 Limestone beds in the Thorsberg quarry in southern Sweden where the fossil meteorites were deposited in the Ordovician ~480 Myr ago³. Total sedimentation time was ~1–2 Myr, estimated by conodont biochronology and average sedimentation rates^{9,3,22}. The number of meteorites analysed per quarry bed is indicated. Each meteorite had a size of between 2 and 9 cm. Reddish-brown limestones are shown shaded dark, grey sediments are shaded light grey.

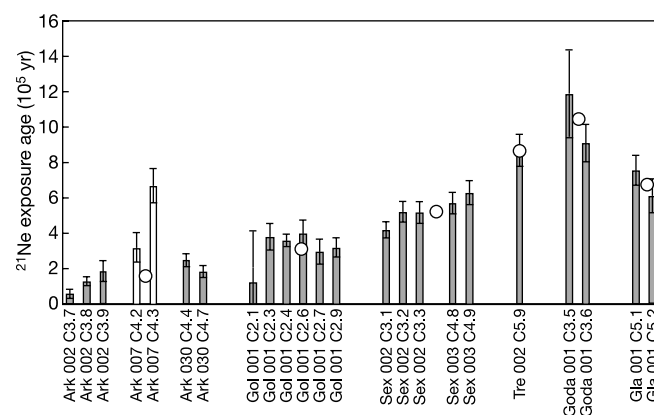


Figure 2 Distribution of ²¹Ne exposure ages. Meteorites from oldest sediments are shown on the left-hand side. Ark 007 samples have large nucleogenic neon corrections and are therefore discarded (Supplementary Discussion). Ages of Gla 001 may be too low owing to unusually large meteorite size or moderate losses of cosmogenic Ne (see main text). Error bars are 1σ, and include weighing uncertainties, ion statistics, and blanks. Open circles are averaged ²¹Ne exposure ages of all analyses from the same quarry bed. Ark, Arkeologen; Gol, Golvsten; Gla, Glaskarten; Sex, Sextummen; Tre, Tredje Karten; Goda, Goda Lagret.

All grain batches have retained cosmogenic ^3He and ^{21}Ne acquired during the journey of their parent meteorite to Earth (minor corrections for non-cosmogenic contributions on these isotopes are explained in Methods and Supplementary Discussion). Table 1 also shows nominal cosmic-ray exposure ages (the time a meteorite spends as a small object in space) deduced from both ^3He and ^{21}Ne . For all six samples of Golvsten 001 (Gol 001), ^{21}Ne ages as well as ^3He ages agree within uncertainties, and the mean ^3He age agrees with the mean ^{21}Ne age within the estimated error of the production rate ratio $P(^3\text{He})/P(^{21}\text{Ne})$ in chromite (Methods). This indicates that these grains cannot have lost a considerable fraction of their ^3He , in agreement with their excellent state of preservation (Supplementary Fig. 1). However, the less well-preserved grains from most other meteorites must have lost a major fraction of their ^3He , as nominal ^3He ages in Table 1 are often considerably below the respective ^{21}Ne ages. We therefore base our arguments on the ^{21}Ne ages only. These are shown in Fig. 2.

The ^{21}Ne ages of different samples from the same meteorite always agree within error limits, providing evidence that the chromite grains have been retentive. Figure 2 shows three additional notable features. First, excluding Ark 007, there is agreement between the mean exposure ages of samples from different meteorites from the same bed, confirming the integrity of the data. Second, exposure ages are all close to or less than a million years. This is at the very low end of ages typical for L chondrites falling at present, which mostly range between ~ 3 and ~ 60 Myr (refs 10, 11). Third, exposure ages generally increase according to their position in the sediment column, and the difference between the highest and lowest ages is consistent with the total deposition time of the sediments of 1–2 Myr. The only exceptions are Ark 007 and Gla 001, discussed below.

These observations provide strong additional support for the conclusion³ that all fossil meteorites derive from one large collision in the asteroid belt. However, the low nominal exposure ages first require critical assessment of their validity. Ages that are too low could be the result of overestimated production rates or losses of cosmogenic Ne. The overall data pattern is inconsistent with both possibilities. The first would require that all fossil meteorites were interior fragments from unusually large pre-atmospheric bodies of perhaps several metres size. The age progression in Fig. 2 and the

similar nominal ages of different meteorites from the same bed would then be a pure coincidence, because production rates in very large bodies strongly depend on the sample position. The regular data pattern is also inconsistent with very substantial gas losses of the order of 90% or more—such losses are required if the true exposure ages of the fossil meteorites were in the range of those of most L chondrites falling today.

As noted, two meteorites do not perfectly fit the overall pattern. The two analyses of Ark 007 yield higher nominal exposure ages than all five samples from two other meteorites from the same bed, and these analyses also display the largest relative age difference of any two samples from the same meteorite. As explained in the Supplementary Discussion, the most likely reason for this is a large contribution of nucleogenic ^{21}Ne (from the reaction $^{18}\text{O}(\alpha, n)^{21}\text{Ne}$) in these two samples which show the highest ^4He concentrations, that is, the highest fluence of α -particles. Therefore, the Ark 007 results can reasonably be dismissed. The nominal exposure age of sample Gla 001 is somewhat lower than would be expected from its position at the top of the studied sediment column. A larger than average meteorite size is a viable explanation in this case, as argued in the Supplementary Discussion, although ^{21}Ne loss of the order of 40–50% cannot firmly be excluded. Note that it is very unlikely that either Ark 007 or Gla 001 are unrelated to the collision that produced the other meteorites, for two reasons: first, their atypically low exposure ages compared to modern falls would then be a coincidence; and second, it is a priori unlikely that we sampled a meteorite from the normal background population, which represents only $\sim 1\%$ of the fossil meteorites.

We conclude that the low nominal ^{21}Ne exposure ages in Fig. 2 are essentially correct. Losses of ^{21}Ne —if any—must be minor. The exposure age progression of meteorites from bottom to top of the sediment column, which inversely matches the relative sediment ages to within better than a factor of two, yields strong independent support for the hypothesis³ that the fossil meteorites are all derived from one single very large asteroid collision. This event was probably the same event that reset the K-Ar clocks of many L chondrites some 500 Myr ago. Therefore, the chromite grains in the fossil meteorites allow us to deduce transport times to Earth of fragments from a large asteroid collision. This work demonstrates

Table 1 Helium and neon concentrations and ^3He (T_3) and ^{21}Ne (T_{21}) exposure ages

Sample	^3He ($10^{-10} \text{ cm}^3 \text{ g}^{-1}$)	^4He ($10^{-5} \text{ cm}^3 \text{ g}^{-1}$)	$^{21}\text{Ne}_{\text{cos}}$ ($10^{-10} \text{ cm}^3 \text{ g}^{-1}$)	T_3 (10^5 yr)	T_{21} (10^5 yr)
Gla 001 C5.1	44.1 $\pm$ 4.8	32.4 $\pm$ 3.2	3.78 $\pm$ 0.39	3.55 $\pm$ 0.15	7.51 $\pm$ 0.83
Gla 001 C5.2	34.3 $\pm$ 3.8	27.6 $\pm$ 2.8	3.00 $\pm$ 0.44	2.71 $\pm$ 0.13	6.07 $\pm$ 0.94
Goda 001 C3.5	23.1 $\pm$ 4.7	35.4 $\pm$ 7.1	5.54 $\pm$ 1.2	1.89 $\pm$ 0.38	11.8 $\pm$ 2.5
Goda 001 C3.6	15.2 $\pm$ 1.5	49.8 $\pm$ 5.0	4.23 $\pm$ 0.49	1.24 $\pm$ 0.12	9.06 $\pm$ 1.0
Tre 002 C5.9	39.3 $\pm$ 4.2	8.42 $\pm$ 0.84	6.21 $\pm$ 0.65	3.15 $\pm$ 0.19	8.64 $\pm$ 0.9
Sex 002 C3.1	46.5 $\pm$ 5.0	2.53 $\pm$ 0.25	1.98 $\pm$ 0.21	3.79 $\pm$ 0.40	4.14 $\pm$ 0.44
Sex 002 C3.2	40.3 $\pm$ 4.3	3.40 $\pm$ 0.34	2.47 $\pm$ 0.27	3.28 $\pm$ 0.35	5.16 $\pm$ 0.57
Sex 002 C3.3	28.4 $\pm$ 3.2	1.49 $\pm$ 0.15	2.46 $\pm$ 0.29	2.32 $\pm$ 0.26	5.13 $\pm$ 0.61
Sex 003 C4.8	33.3 $\pm$ 3.3	3.17 $\pm$ 0.32	2.74 $\pm$ 0.29	2.71 $\pm$ 0.27	5.67 $\pm$ 0.60
Sex 003 C4.9	34.5 $\pm$ 3.5	7.87 $\pm$ 0.79	3.01 $\pm$ 0.32	2.82 $\pm$ 0.28	6.24 $\pm$ 0.67
Gol 001 C2.1	23.8 $\pm$ 2.1	5.37 $\pm$ 0.41	0.58 $\pm$ 1.4	1.94 $\pm$ 0.17	1.19 $\pm$ 2.9
Gol 001 C2.3	30.6 $\pm$ 3.5	3.83 $\pm$ 0.38	1.83 $\pm$ 0.36	2.49 $\pm$ 0.29	3.75 $\pm$ 0.73
Gol 001 C2.4	37.5 $\pm$ 1.5	3.94 $\pm$ 0.06	1.73 $\pm$ 0.17	3.05 $\pm$ 0.12	3.55 $\pm$ 0.34
Gol 001 C2.6	35.4 $\pm$ 5.4	8.03 $\pm$ 1.2	1.93 $\pm$ 0.36	2.89 $\pm$ 0.44	3.95 $\pm$ 0.74
Gol 001 C2.7	29.2 $\pm$ 6.4	6.33 $\pm$ 0.8	1.43 $\pm$ 0.34	2.38 $\pm$ 0.52	2.92 $\pm$ 0.70
Gol 001 C2.9	34.3 $\pm$ 4.5	4.97 $\pm$ 0.6	1.53 $\pm$ 0.26	2.80 $\pm$ 0.37	3.14 $\pm$ 0.54
Ark 002 C3.7	11.2 $\pm$ 1.2	0.69 $\pm$ 0.07	0.27 $\pm$ 0.05	—	0.53 $\pm$ 0.24
Ark 002 C3.8	22.1 $\pm$ 2.5	1.04 $\pm$ 0.10	0.62 $\pm$ 0.11	—	1.24 $\pm$ 0.24
Ark 002 C3.9	65.1 $\pm$ 7.2	3.24 $\pm$ 0.32	0.91 $\pm$ 0.19	—	1.81 $\pm$ 0.58
Ark 007 C4.2	35.5 $\pm$ 4.8	206 $\pm$ 26	(1.47 $\pm$ 0.39)	(2.90 $\pm$ 0.39)	(3.14 $\pm$ 0.83)
Ark 007 C4.3	50.8 $\pm$ 7.7	55.81 $\pm$ 7.9	(3.10 $\pm$ 0.45)	(4.14 $\pm$ 0.63)	(6.65 $\pm$ 0.97)
Ark 030 C4.4	11.6 $\pm$ 1.3	3.52 $\pm$ 0.36	1.15 $\pm$ 0.18	0.94 $\pm$ 0.10	2.45 $\pm$ 0.38
Ark 030 C4.7	9.12 $\pm$ 1.7	3.04 $\pm$ 0.54	0.83 $\pm$ 0.15	0.74 $\pm$ 0.14	1.79 $\pm$ 0.33

Data for 23 chromite grain batches from 9 fossil meteorites. All ^3He assumed to be cosmogenic (except for Ark 002 and Ark 007, Supplementary Discussion). Cosmogenic ^{21}Ne ($^{21}\text{Ne}_{\text{cos}}$) calculated by subtracting trapped and nucleogenic Ne (Methods). Production rates used to calculate ^3He and ^{21}Ne exposure ages (T_3 , T_{21}) explained in Methods. ^3He and ^{21}Ne ages of Ark 007 in parentheses owing to large contributions by nucleogenic components. ^3He ages for Ark 002 not given owing to solar ^3He (Supplementary Discussion). Uncertainties (1 σ) include weighing errors, ion statistics, and blank corrections. Complete noble gas data given in Supplementary Table 1.

that the first fragments from the L-chondrite parent body break-up event arrived on Earth only 100,000 to 200,000 yr after the impact (the ^{21}Ne ages of the Arkeologen meteorites). The low exposure ages of the Arkeologen meteorites are also consistent with the conclusion that the asteroid disruption occurred shortly before the deposition of the Arkeologen bed⁶. This conclusion is based on the distribution of extraterrestrial chromite grains dispersed in sediment layers.

Numerical simulations of asteroid and meteorite delivery from the main asteroid belt into the inner Solar System predict that a body which has moved into an orbit in resonance with that of a planet or the Solar System's natural frequencies typically will collide with the Sun (or rarely a terrestrial planet) within a very few million years^{1,12}. The fact that exposure ages of stony meteorites typically are an order of magnitude longer is explained by a slow drift of a meteorite freshly ejected from its parent body into a resonance by the so-called Yarkovsky effect, which is a momentum transfer due to re-radiation of solar energy at infrared wavelengths¹³. Hence, very short total exposure ages are only to be expected if meteorites are directly injected into an orbital resonance from a nearby impact. The most probable candidates are very large collisions creating families of asteroids with similar orbital elements^{1,2}.

The exposure ages of the fossil meteorites confirm the dynamical models to the extent that they show that at least one strong orbital resonance in the main asteroid belt is capable of delivering material to Earth within the short timescales suggested by the models. The data also show that an L chondrite parent body was located close enough to a powerful resonance that collisional ejecta could directly reach a resonant orbit. The two most prolific resonances are situated in the inner main belt^{14,15}. This is consistent with the hypothesis that L chondrites and other ordinary chondrites come from the S(IV) subset of the S asteroids, the most abundant spectroscopic class in the inner belt¹⁸. The fossil meteorites in southern Sweden thus probably originate in the inner main asteroid belt. A viable parent body is the precursor of the Flora family near 2.2 astronomical units from the Sun¹⁹. □

Methods

Samples and analytical methods

We analysed a total of 23 batches of one to a few chromite grains (average grain size 80–100 μm ; mass of each batch 4–40 μg)³. Chromite grains were separated from bulk fossil meteorites³. Major element composition was determined using SEM/EDAX. Most of the grains studied have compositions very similar to the average composition of chromites from 26 fossil meteorites analysed by Schmitz *et al.*³. Only the grains from Tre 002 have about twice as much MgO (~5%) as chromites from the other meteorites.

Gases were extracted by melting grains with a Nd-YAG CW laser. Samples were heated for 3–4 min. Gases were cleaned with several liquid nitrogen cold traps and ZrAl 'getters' to remove active gases. Helium and Ne were separated before analysis by a cryotrap in order to measure He and Ne isotopes with the appropriate mass resolution. The mass spectrometer is equipped with a molecular drag pump compressing the sample gas into the ion-source, resulting for He and Ne in a sensitivity increase of about two orders of magnitude compared to conventional noble gas mass spectrometers⁸. The sensitivity of the mass spectrometer is about $7.2 \times 10^{14} \text{ Hz cm}^{-3} \text{ STP}$ for He and $2.8 \times 10^{15} \text{ Hz cm}^{-3} \text{ STP}$ for Ne, respectively, yielding count rates for samples of the order of 10–100 Hz for ^3He and ^{21}Ne . Spectrometer noble gas memory was monitored before and after gas inlet, and signals were corrected accordingly. Gas amounts were calibrated using gas from a standard air reservoir, believed to be known to better than 3%.

Production rates

Because no information on pre-atmospheric meteorite size and sample depth can be deduced from our data, we have to assume nuclide production rates typical for 'average-sized' meteorites. Elemental production rates²⁰ of cosmogenic ^3He and ^{21}Ne by galactic cosmic rays were calculated assuming a typical meteorite radius of 25 cm and a sample position about 10 cm below the surface. Since the atomic mass difference between Fe and Ni, and between Cr and Fe, both are ~2 mass units, we assume the production rate differences between $P_{21\text{Ne}}(\text{Cr})$ and $P_{21\text{Ne}}(\text{Fe})$ to be similar to the difference between $P_{21\text{Ne}}(\text{Fe})$ and $P_{21\text{Ne}}(\text{Ni})$. Overall uncertainties of production rates are estimated to be about 40%, including uncertainties due to unknown meteorite sizes and sample positions (unless pre-atmospheric sizes of the fossil meteorites would have been considerably larger than typical values of present day meteorites, see main text).

Relevant target elements for cosmogenic ^3He production are O, Cr, Fe, Al and Mg (in the order of decreasing contribution). Cosmogenic ^{21}Ne is mainly produced by spallation on Mg, Al, Cr and Fe. Although Cr and Fe are the major compositional constituents of

chromite besides oxygen, much of the total ^{21}Ne production is from Mg and Al (on average 43% from Mg, 34% from Al, 16% from Cr and 7% from Fe).

Production rates (calculated from elemental production rates²⁰ and major element composition) are on average $1.23 \times 10^{-8} \text{ cm}^3 \text{ STP g}^{-1} \text{ Myr}^{-1}$ for ^3He and in the range $(4.7\text{--}7.2) \times 10^{-10} \text{ cm}^3 \text{ STP g}^{-1} \text{ Myr}^{-1}$ for ^{21}Ne .

Corrections for non-cosmogenic noble gases

Trapped Ne has been corrected for by assuming atmospheric composition for all samples except Ark 002, for which solar wind composition was assumed (Supplementary Fig. 2). Nucleogenic ^{21}Ne produced when ^{18}O captures α -particles from U/Th decay was calculated by assuming all ^4He to be radiogenic and a ratio²¹ of

$$^{21}\text{Ne}_{\text{nuc}}/^4\text{He}_{\text{rad}} = 2.8 \times 10^{-8} \text{ for a mean O concentration of the chromites of 33.6\% (ref. 3). Corrections for } ^{21}\text{Ne}_{\text{nuc}} \text{ are between 0.2\% and 3.3\% only, except for Ark 007 (up to 39\%, Supplementary Discussion). A correction for nucleogenic } ^3\text{He} \text{ produced through the reaction } ^6\text{Li}(n,\alpha)^3\text{H}(\beta^-)^3\text{He} \text{ is not possible, since Li concentrations in the grains are unknown. They are expected to be very low, however. Furthermore, } ^3\text{He}/^4\text{He} \text{ ratios in the chromites are mostly much higher than typical crustal radiogenic production ratios of } 10^{-7}\text{--}10^{-8}, \text{ indicating that nucleogenic } ^3\text{He} \text{ is negligible. We therefore assume that the } ^3\text{He} \text{ in the chromites is entirely cosmogenic. (Two exceptions, Ark 007 and Ark 002, are discussed in the Supplementary Discussion).}$$

In some grains, the ^{21}Ne excesses are several times higher than what would have been produced during 480-Myr exposure on the very surface of the Earth at the present day altitude (a completely unrealistic scenario). Since the meteorites fell into an epicontinental sea and were later shielded by at least several 10 m of sediments essentially until their recovery, cosmogenic nuclide production on Earth fails by many orders of magnitude to explain the observed excesses.

Received 19 April; accepted 7 June 2004; doi:10.1038/nature02736.

- Gladman, B. J. *et al.* Dynamical lifetimes of objects injected into asteroid belt resonances. *Science* **277**, 197–201 (1997).
- Zappalà, V., Cellino, A., Gladman, B. J., Manley, S. & Migliorini, F. Asteroid showers on Earth after family breakup events. *Icarus* **134**, 176–179 (1998).
- Schmitz, B., Tassinari, M. & Peucker-Ehrenbrink, B. A rain of ordinary chondritic meteorites in the early Ordovician. *Earth Planet. Sci. Lett.* **194**, 1–15 (2001).
- Keil, K., Haack, H. & Scott, E. R. D. Catastrophic fragmentation of asteroids: Evidence from meteorites. *Planet. Space Sci.* **42**, 1109–1122 (1994).
- Bogard, D. D. Impact ages of meteorites: A synthesis. *Meteoritics* **30**, 244–268 (1995).
- Schmitz, B., Häggström, T. & Tassinari, M. Sediment-dispersed extraterrestrial chromite traces a major asteroid disruption event. *Science* **300**, 961–964 (2003).
- Patterson, D. B., Farley, K. A. & Schmitz, B. Preservation of extraterrestrial ^3He in 480-Ma-old marine limestones. *Earth Planet. Sci. Lett.* **163**, 315–325 (1998).
- Baur, H. A noble-gas mass spectrometer compressor source with two orders of magnitude improvement in sensitivity. *Eos* **46**, F1118 (1999).
- Schmitz, B., Lindström, M., Asaro, F. & Tassinari, M. Geochemistry of meteorite-rich marine limestone strata and fossil meteorites from the lower Ordovician at Kinnekulle, Sweden. *Earth Planet. Sci. Lett.* **145**, 31–48 (1996).
- Marti, K. & Graf, T. Cosmic-ray exposure history of ordinary chondrites. *Annu. Rev. Earth Planet. Sci.* **20**, 221–243 (1992).
- Wieler, R. Cosmic-ray produced noble gases in meteorites. *Rev. Mineral. Geochem.* **47**, 125–170 (2002).
- Farinella, P., Vokrouhlický, D. & Morbidelli, A. in *Accretion of Extraterrestrial Matter Throughout Earth's History* (eds Peucker-Ehrenbrink, B. & Schmitz, B.) 31–49 (Kluwer, New York, 2001).
- Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Brož, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 395–408 (Univ. Arizona Press, Tucson, 2002).
- Morbidelli, A. & Gladman, B. Orbital and temporal distributions of meteorites originating in the asteroid belt. *Meteorit. Planet. Sci.* **33**, 999–1016 (1998).
- Bottke, W. F. *et al.* Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
- Gaffey, M. J. *et al.* Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993).
- Binzel, R. P., Lupishko, D. F., Di Martino, M., Whiteley, R. J. & Hahn, G. J. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 255–271 (Univ. Arizona Press, Tucson, 2002).
- Gradie, J. & Tedesco, E. F. Compositional structure of the asteroid belt. *Science* **216**, 1405–1407 (1983).
- Nesvorný, D., Morbidelli, A., Vokrouhlický, D., Bottke, W. F. & Brož, M. The Flora family: A case of the dynamically dispersed collisional swarm? *Icarus* **157**, 155–172 (2002).
- Leya, I., Lange, H.-J., Neumann, S., Wieler, R. & Michel, R. The production of cosmogenic nuclides in stony meteoroids by galactic cosmic-ray particles. *Meteorit. Planet. Sci.* **35**, 259–286 (2000).
- Leya, I. & Wieler, R. Nucleogenic production of Ne isotopes in Earth's crust and upper mantle induced by alpha particles from the decay of U and Th. *J. Geophys. Res.* **104**, 15439–15450 (1999).
- Jianhua, Z. Middle Ordovician conodonts from the Atlantic faunal region and the evolution of key conodont genera. *Meddn Stockh. Univ. Inst. Geol. Geokem.* **298**, 1–27 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Tassinari for help with recovery of chromite grains, I. Leya for discussions, G. Quitté and U. Wiechert for help in the laboratory and W. F. Bottke for comments on the manuscript. Field work was funded by the National Geographic Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Ph.R.H. (heck@erdw.ethz.ch).

A substantial amount of hidden magnetic energy in the quiet Sun

J. Trujillo Bueno^{1,2}, N. Shchukina³ & A. Asensio Ramos¹

¹Instituto de Astrofísica de Canarias, E-38205 La Laguna, Tenerife, Spain

²Consejo Superior de Investigaciones Científicas, E-28006 Madrid, Spain

³Main Astronomical Observatory, National Academy of Sciences, Zabolotnogo 27, 03680 Kyiv, Ukraine

Deciphering and understanding the small-scale magnetic activity of the quiet solar photosphere should help to solve many of the key problems of solar and stellar physics, such as the magnetic coupling to the outer atmosphere and the coronal heating^{1–3}. At present, we can see only ~1 per cent of the complex magnetism of the quiet Sun^{1,4–7}, which highlights the need to develop a reliable way to investigate the remaining 99 per cent. Here we report three-dimensional radiative transfer modelling of scattering polarization in atomic and molecular lines that indicates the presence of hidden, mixed-polarity fields on subresolution scales. Combining this modelling with recent observational data^{8–11}, we find a ubiquitous tangled magnetic field with an average strength of ~130 G, which is much stronger in the intergranular regions of solar surface convection than in the granular regions. So the average magnetic energy density in the quiet solar photosphere is at least two orders of magnitude greater than that derived from simplistic one-dimensional investigations^{12,13}, and sufficient to balance radiative energy losses from the solar chromosphere.

Most of our present empirical knowledge of solar surface magnetism stems from the analysis of the light polarization that the Zeeman effect induces in spectral lines^{1,4–7}. For example, the circular polarization signals of which synoptic magnetograms are made reveal the existence of an irregular network of spatially unresolved, intermittent flux patches of kilogauss field concentrations outlining the boundaries of the giant velocity cells of the supergranulation¹. Such routine magnetograms give the (wrong) impression that the cell interiors (referred to here as internetwork regions) are non-magnetic, and it is important to note that such regions cover most of the solar ‘surface’ at any given time during the solar magnetic activity cycle. However, high-spatial-resolution magnetograms show a multitude of mixed magnetic polarities within the internetwork regions, with most of the detected flux located in the intergranular lanes of the solar granulation pattern where the plasma is downflowing^{4–7}. The mean unsigned flux density turns out to be ~10 G when choosing the best compromise between polarimetric sensitivity and spatio-temporal resolution that is at present feasible. When such polarization signals are interpreted taking into account the fact that the magnetic field is not being spatially resolved (for example, by assuming that one or more magnetic components are coexisting with a ‘non-magnetic’ component within the spatio-temporal resolution element of the observation), it is then found that the filling factor of the magnetic component(s) is ~1%. The ‘problem’ with the Zeeman effect is that the amplitudes of the measured polarization signals are smaller the greater the degree of cancellation of mixed magnetic polarities within the spatio-temporal resolution element of the observation. Therefore, lack of detection does not necessarily imply absence of magnetic fields.

Fortunately, scattering processes in spectral lines produce linear polarization signals, whose amplitudes are efficiently modified in the presence of tangled magnetic fields of strength $B_H \approx 1.137 \times 10^{-7} / (t_{\text{life}} g_L)$ (with B_H expressed in gauss, and where t_{life} and g_L are respectively the lifetime in seconds and Landé factor of the upper level of the transition). This so-called Hanle effect^{14,15} has the required diagnostic potential for investigating spatially unresolved, ‘hidden’ mixed-polarity magnetic fields

in the solar atmosphere¹⁶. The problem is how to apply it to obtain reliable information, given that Hanle-effect diagnostics of such ‘turbulent’ fields rely on a comparison between the observed scattering polarization and that corresponding to the zero-field reference case. It has been pointed out correctly^{12,13} that a suitable spectral line for Hanle-effect diagnostics of ‘turbulent’ photospheric fields is that of Sr I at 4,607 Å. However, the simplified approach of assuming that the highly inhomogeneous and dynamic solar photosphere can be well represented by a one-dimensional (1D) and static model is very unreliable, because of the need to use the free parameters of ‘classical’ stellar spectroscopy (that is, micro- and macroturbulence for line broadening), which have a serious impact on the calculated polarization amplitudes¹⁷. We have shown¹⁷ that the particular 1D approach that has been applied^{12,13} yields artificially low values for the strength of the ‘turbulent’ field—that is, between 20 and 10 G as shown by the dashed lines of Fig. 3 in ref. 17.

In order to improve the reliability of diagnostic tools based on the Hanle effect, we have developed a novel approach based on multi-level scattering polarization calculations in three-dimensional (3D) models of the solar photosphere, which we have obtained from realistic hydrodynamical simulations of solar surface convection¹⁸. This has allowed us to obtain the linear polarization amplitudes that scattering processes in the inhomogeneous solar photosphere would produce in the Sr I 4,607 Å line if there were no magnetic

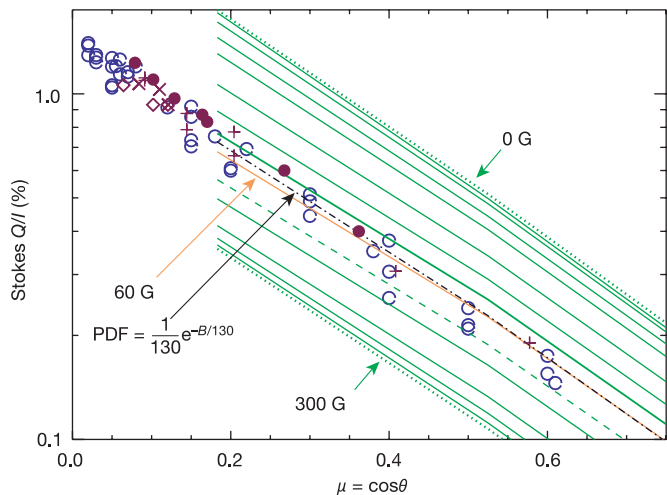


Figure 1 Spectropolarimetric observations versus 3D modelling of the Hanle effect. This figure shows the centre-to-limb variation of the fractional linear polarization at the core of the Sr I 4,607 Å line after subtraction of the continuum polarization level. (The Stokes Q and I parameters are defined in ref. 1; $\mu = \cos \theta$, with θ the angle between the solar radius vector through the observed point and the line of sight.) Open circles, various observations⁸ taken during a minimum period of the solar cycle (in particular, during September–October 1995). The remaining symbols correspond to observations taken during the most recent maximum period of the solar cycle. Diamonds, observations⁹ obtained during May 2000; crosses and ‘plus’ symbols, observations taken during August 2000 (see ref. 10) and December 2002 (V. Bommier; personal communication), respectively; filled circles, observations obtained during September 2003 at the Istituto Ricerche Solari Locarno (Switzerland) in collaboration with M. Bianda. Coloured lines, the results of our 3D scattering polarization calculations in the presence of a volume-filling and single-valued microturbulent field with an isotropic distribution of directions below the mean free path of the line photons (from top to bottom: 0, 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, 150, 200, 250 and 300 G). We have solved the relevant equations²⁸ via the application of efficient radiative transfer methods²⁹, and using realistic collisional depolarizing rate values¹², which turn out to be the largest rates among those found in the literature. Note that there is no evidence of a serious modulation of the strength of the ‘turbulent’ field with the last solar activity cycle, and that the best average fit to the observations is obtained for 60 G. The black dashed-dotted line indicates the resulting Q/I line-core amplitudes for the case of a single exponential PDF ($e^{-B/B_0} / B_0$) with $B_0 = 130$ G, which implies $\langle B \rangle = 130$ G.

field. We point out that our synthetic intensity profiles (which take fully into account the Doppler shifts of the convective flow velocities in the 3D model) are automatically in excellent agreement with the observations when the meteoritic strontium abundance is chosen. However, we find that the calculated fractional linear polarization is substantially larger than the observed one, thus indicating the need to invoke magnetic depolarization.

As seen in Fig. 1, our radiative transfer simulations of the Hanle effect in the Sr I line show that a volume-filling, microturbulent magnetic field of about 60 G leads to a notable agreement with the observed Q/I (where Q and I are Stokes parameters). In Fig. 1 there

is a clear indication that the strength of the ‘turbulent’ field required to explain the Q/I observations decreases with height in the atmosphere, from the 70 G needed to explain the observations at $\mu = 0.6$ to the 50 G required to fit the observations at $\mu = 0.1$ (μ is defined in Fig. 1 legend). This corresponds approximately to a height range between 200 and 400 km above the solar visible ‘surface’.

If the hidden magnetic field of the quiet solar photosphere really had a strength of 60 G at all points in the solar photosphere, then the mean magnetic energy density ($E_m = \langle B^2 \rangle / 8\pi$) would be $E_m \approx 140 \text{ erg cm}^{-3}$, which is an order of magnitude smaller than that corresponding to the kilogauss fields of the network patches.

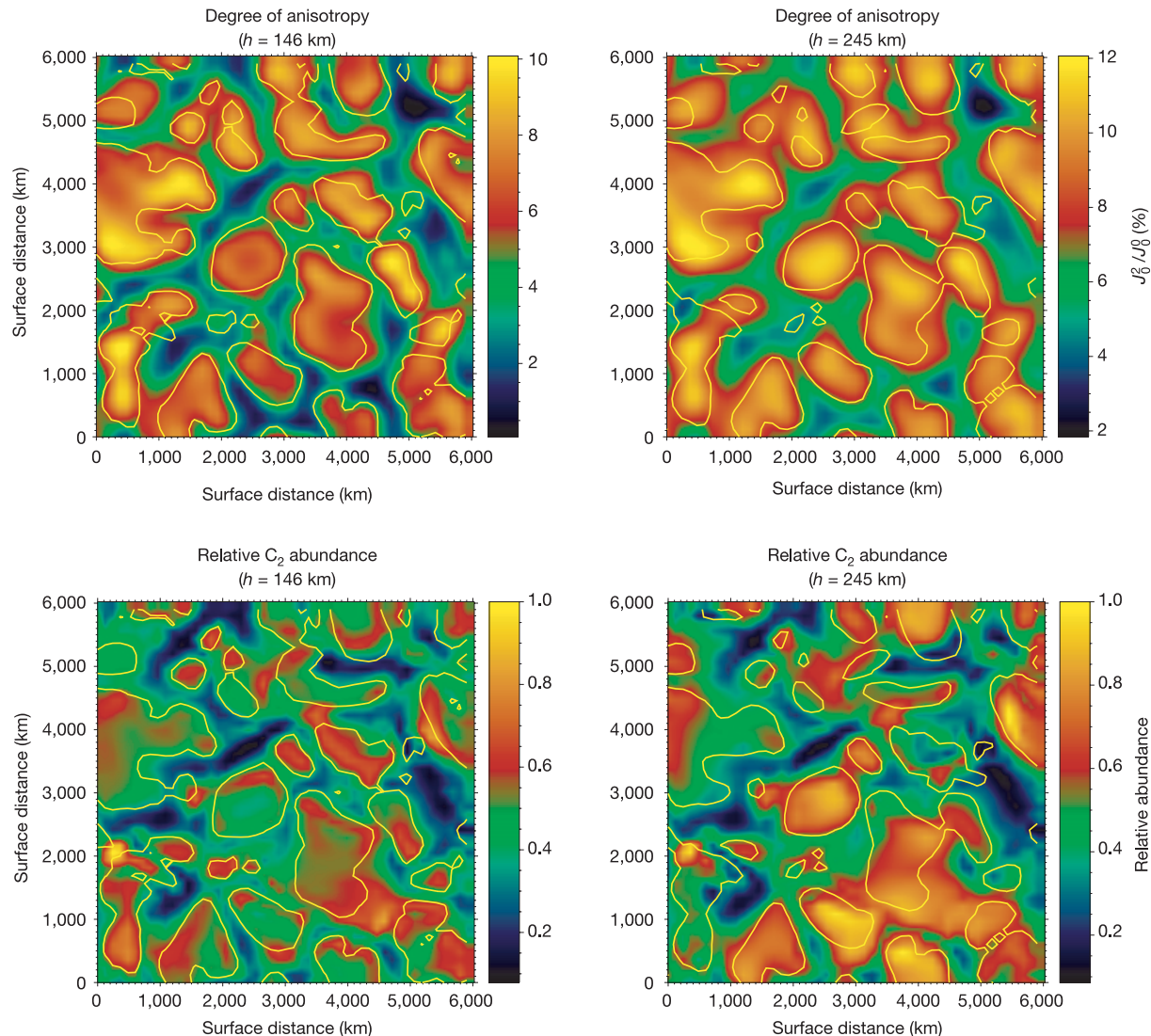


Figure 2 Evidence for a ‘turbulent’ field organized at the spatial scales of the solar granulation pattern. This figure demonstrates that the scattering polarization observed in molecular lines¹¹ comes mainly from the upflowing regions of the ‘quiet’ solar photosphere. The solid line contours delineate such upflowing regions at two heights (h) in the 3D photospheric model, which indicate the approximate atmospheric region where the observed C_2 line radiation originates. The two top panels show the horizontal fluctuation of the ‘degree of anisotropy’ ($\mathcal{A} = J_p^2/J_0^2$; refs 15, 28) of the solar continuum radiation at 5,000 Å. (Note that there is a very strong correlation between the upflowing regions and $\mathcal{A}$.) The two bottom panels visualize the corresponding horizontal fluctuation in the number density ($\mathcal{N}$) of C_2 molecules. (Note that there is a significant correlation between the upflowing regions and $\mathcal{N}$.) As in the solar atmosphere the scattering polarization in weak molecular lines is proportional to both $\mathcal{A}$ and $\mathcal{N}/\eta_c$ (where η_c is the

background continuum opacity, which is anticorrelated with $\mathcal{N}$ in such a range of heights) we conclude that only the upflowing regions make a significant contribution to the observed molecular scattering polarization. Therefore, we can obtain information on the distribution of magnetic fields in such (granular) upflowing regions of the solar photosphere via theoretical modelling of the observed scattering polarization in suitably chosen pairs of C_2 lines³⁰. We have calculated the magnetic sensitivity of the observed C_2 lines by taking into account all the relevant optical pumping mechanisms in a very realistic multilevel model for C_2 . Thanks to this Hanle-effect investigation, we have been able to conclude that in the granular regions the ‘turbulent’ field is much weaker than what is needed to explain the observed depolarization in the Sr I 4,607 Å line. For this reason, the intergranular regions must be pervaded by relatively strong tangled fields capable of producing most of the observed depolarization in the strontium line.

However, both magnetohydrodynamical simulations^{19,20} and observations of the Zeeman effect^{4,21,7} indicate that the photospheric plasma of the quiet Sun has a continuous distribution of field strengths, in such a way that the weaker the field the larger the probability of finding a magnetic strength between B and $B + dB$ when no distinction is made between granular and intergranular regions. As shown by the black dashed-dotted line of Fig. 1, if we assume that the probability distribution function (PDF) has an exponential shape ($e^{-B/B_0}/B_0$), we then find that $B_0 \approx 130$ G yields a fairly good fit to the observed fractional linear polarization. In this much more realistic case $E_m \approx 1,300 \text{ erg cm}^{-3}$ (that is, $\sqrt{\langle B^2 \rangle} \approx 180$ G), which is about 20% of the averaged kinetic energy density produced by convective motions at a height of 200 km in the 3D photospheric model. This result and the fact that the observed Q/I does not seem to be modulated by the solar cycle (see Fig. 1) suggests that a small-scale dynamo associated with turbulent motions within a given convective domain of ionized gas^{19,20} plays a significant role for the quiet Sun magnetism. The total magnetic energy stored in the internetwork regions is now slightly larger than that corresponding to the kilogauss fields of the network patches. It is also of interest to point out that our empirical PDF with $\langle B \rangle \approx 130$ G produces a negligible Zeeman broadening of the emergent spectral line profiles, which is fully compatible with the reported empirical constraints¹. However, in contrast to what might be expected from the conclusions of a recent paper²², the stretched exponential PDF that characterizes the surface distribution of magnetic fields 'predicted' by the most recent numerical experiments of turbulent dynamos¹⁹, which correspond to a magnetic Reynolds number of about 1,000, implies a Hanle-effect reduction of the scattering polarization amplitude of the Sr I 4,607 Å line that is significantly smaller than what is needed to explain the observations.

We do not have a sufficient number of observational constraints to be able to determine the exact shape or the detailed spatial variation of the solar PDF. However, we can conclude that the 'turbulent' field is organized at the spatial scales of the solar granulation pattern, with relatively weak fields above the granules and with much stronger fields above the intergranular lanes (Fig. 2). As shown in Fig. 2 legend, most of the volume of the upflowing granules where the observed C_2 line polarization originates is occupied by magnetic fields much weaker than what is needed to explain the observed depolarization in the Sr I 4,607 Å line. Therefore, most (but not all) of the observed strontium line depolarization must be produced by relatively strong and tangled fields in the intergranular regions.

The contribution of the intergranular plasma to the observed scattering polarization in the strontium line is very likely to be close to the regime of Hanle saturation, which for this spectral line occurs for magnetic strengths $B \gtrsim 300$ G (Fig. 1). This helps us to understand why, with low spatial and temporal resolution, it is not easy (but definitely not impossible) to detect spatial fluctuations in the observed Q/I (Fig. 1). We have carried out calculations with several plausible PDFs, assuming in all cases that the angular distribution of the field vectors is isotropic and microturbulent for every single field strength. For instance, we have assumed an exponential PDF ($\text{PDF}_G = e^{-B/B_g}/B_g$) for the upflowing regions with a free parameter B_g to be determined from the observations and a given maxwellian function ($\text{PDF}_I = 2.38 \times 10^{-8} B^2 \exp(-B/456)^2$) for the downflowing plasma obtained from a best fit to the strong field part of the intergranular histogram of the Zeeman splittings observed in near-infrared lines of neutral iron⁷. Note that this maxwellian PDF implies that the filling factor of kilogauss fields is $\sim 2\%$ —that is, it does not exclude the possibility⁶ of small-scale kilogauss fields in the internetwork regions. With this 'educated' choice of plausible PDFs, the best agreement between theory and observations is obtained for $B_g \approx 15$ G, which is consistent with the results of our investigation of the Hanle effect in C_2 lines. We find

that most of the magnetic energy turns out to be due to rather chaotic fields in the intergranular plasma with strengths between the equipartition field values and ~ 1 kG. The total magnetic energy stored in the internetwork regions is now substantially larger than that corresponding to the kilogauss fields of the supergranulation network.

Our conclusion that most of the volume of the granular regions is occupied by very weak fields seems to be compatible with the constraints imposed by our present understanding²³ of the 'enigmatic' scattering polarization observed in the Na I D_1 line, whose line-core radiation originates a few hundred kilometres higher in the solar atmosphere. As demonstrated in ref. 23, the linear polarization signature of the scattered light in the D_1 line is only completely suppressed in the presence of magnetic fields larger than 10 G, regardless of their inclination. Note that this result is significantly different to what had been concluded in an earlier paper²⁴. In any case, it is important to point out that the observed linear polarization in the sodium D_1 line still remains enigmatic because nobody has yet been able to model both the amplitude and shape of the observed Q/I profile²⁵.

Our empirical findings may have far-reaching implications in solar and stellar physics. The hot outer regions of the solar atmosphere (chromosphere and corona) radiate and expand, which takes energy. By far the largest energy losses stem from chromospheric radiation, with a total energy flux of $\sim 10^7 \text{ erg cm}^{-2} \text{ s}^{-1}$. Considering our most conservative estimate for the magnetic energy density — that is, $E_m \approx 140 \text{ erg cm}^{-3}$ — we obtain an energy flux similar to the above-mentioned chromospheric energy losses when using either the typical value of $\sim 1 \text{ km s}^{-1}$ for the convective velocities or the Alfvén speed ($v_A = B/(4\pi\rho)^{1/2}$, with ρ the gas density). In reality, as pointed out above, the true magnetic energy density that at any given time during the solar cycle is stored in the quiet solar photosphere is very much larger than 140 erg cm^{-3} . Only a relatively small fraction would thus suffice to balance the energy losses of the solar outer atmosphere.

Equally interesting is our result that $\langle B \rangle \approx 130$ G, which indicates that the unsigned magnetic flux density in the quiet solar photosphere is substantially large. In fact, it has been suggested recently that the dynamic geometry of the magnetic connection between the photosphere and the corona may sensitively depend on the amount of magnetic flux that exists in the internetwork regions³. The hidden flux that we have diagnosed turns out to be organized at the spatial scales of the solar granulation pattern, with much stronger fields in the intergranular regions. Therefore, if the magnetic field of the 'quiet' solar chromosphere is found to be more complex than simply canopy-like, because of the influence of the intergranular magnetic fields which are continuously moving around on temporal scales of minutes, then the possibilities for magnetic reconnection²⁷ and energy dissipation in the solar outer atmosphere would be much enhanced. The 'hidden' magnetic fields that we have reported here could thus provide the clue to understanding how the 10^6 K solar corona is heated. □

Received 13 January; accepted 17 May 2004; doi:10.1038/nature02669.

1. Stenflo, J. O. *Solar Magnetic Fields: Polarized Radiation Diagnostics* (Kluwer, Dordrecht, 1994).
2. Schrijver, C. J. et al. Large-scale coronal heating by the small-scale magnetic field of the Sun. *Nature* **394**, 152–154 (1998).
3. Schrijver, C. J. & Title, A. The magnetic connection between the solar photosphere and the corona. *Astrophys. J.* **597**, L165–L168 (2003).
4. Lin, H. & Rimmele, T. The granular magnetic fields of the quiet Sun. *Astrophys. J.* **514**, 448–455 (1999).
5. Sánchez Almeida, J. & Lites, B. Physical properties of the solar magnetic photosphere under the MISMA hypothesis. II. Network and internetwork fields at the disk center. *Astrophys. J.* **532**, 1215–1229 (2000).
6. Domínguez Cerdeña, I., Kneer, F. & Sánchez Almeida, J. Quiet-Sun magnetic fields at high spatial resolution. *Astrophys. J.* **582**, L55–L58 (2003).
7. Khomenko, E., Collados, M., Solanki, S. K., Lagg, A. & Trujillo Bueno, J. Quiet-Sun internetwork magnetic fields observed in the infrared. *Astron. Astrophys.* **408**, 1115–1135 (2003).
8. Stenflo, J. O., Bianda, M., Keller, C. & Solanki, S. K. Center-to-limb variation of the second solar spectrum. *Astron. Astrophys.* **322**, 985–994 (1997).

9. Trujillo Bueno, J., Collados, M., Paletou, F. & Molodij, G. in *Advanced Solar Polarimetry: Theory, Observations and Instrumentation* (ed. Sigwarth, M.) 141–149 (ASP Conf. Ser. Vol. 236, Astronomical Society of the Pacific, San Francisco, 2001).
10. Bomnier, V. & Molodij, G. Some THEMIS-MTR observations of the second solar spectrum (2000 campaign). *Astron. Astrophys.* **381**, 241–252 (2002).
11. Gandorfer, A. *The Second Solar Spectrum* Vol. 1, 4625 Å to 6995 Å (vdf, Zurich, 2000).
12. Faurobert-Scholl, M., Feautrier, N., Machefer, F., Petrovay, K. & Spielfiedel, A. Turbulent magnetic fields in the solar photosphere: diagnostics and interpretation. *Astron. Astrophys.* **298**, 289–302 (1995).
13. Faurobert, M., Arnaud, J., Vignieu, J. & Frisch, H. Investigation of weak solar magnetic fields. New observational results for the Sr I 460.7 nm linear polarization and radiative transfer modeling. *Astron. Astrophys.* **378**, 627–634 (2001).
14. Hanle, W. Über magnetische Beeinflussung der Polarisation der Resonanzfluoreszenz. *Z. Phys.* **30**, 93–105 (1924).
15. Trujillo Bueno, J. in *Advanced Solar Polarimetry: Theory, Observations and Instrumentation* (ed. Sigwarth, M.) 161–195 (ASP Conf. Ser. Vol. 236, Astronomical Society of the Pacific, San Francisco, 2001).
16. Stenflo, J. O. The Hanle effect and the diagnostics of turbulent magnetic fields in the solar atmosphere. *Sol. Phys.* **80**, 209–226 (1982).
17. Shchukina, N. & Trujillo Bueno, J. in *Solar Polarization 3* (eds Trujillo Bueno, J. & Sánchez Almeida, J.) 336–343 (ASP Conf. Ser. Vol. 307, Astronomical Society of the Pacific, San Francisco, 2003).
18. Asplund, M., Nordlund, Å., Trampedach, R., Allende Prieto, C. & Stein, R. F. Line formation in solar granulation. I. Fe line shapes, shifts and asymmetries. *Astron. Astrophys.* **359**, 729–742 (2000).
19. Cattaneo, F. On the origin of magnetic fields in the quiet photosphere. *Astrophys. J.* **515**, L39–L42 (1999).
20. Stein, R. F. & Nordlund, Å. in *Modelling of Stellar Atmospheres* (eds Piskunov, N. E., Weiss, W. W. & Gray, D. F.), 169–180 (ASP Conf. Ser. Vol. IAU 210, Astronomical Society of the Pacific, San Francisco, 2003).
21. Socas-Navarro, H. & Sánchez Almeida, J. Magnetic fields in the quiet Sun: observational discrepancies and unresolved structure. *Astrophys. J.* **593**, 581–586 (2003).
22. Sánchez Almeida, J., Emonet, T. & Cattaneo, F. Polarization of photospheric lines from turbulent dynamo simulations. *Astrophys. J.* **585**, 536–552 (2003).
23. Trujillo Bueno, J., Casini, R., Landolfi, M. & Landi Degl'Innocenti, E. The physical origin of the scattering polarization of the Na I D lines in the presence of weak magnetic fields. *Astrophys. J.* **566**, L53–L57 (2002).
24. Landi Degl'Innocenti, E. Evidence against turbulent and canopy-like magnetic fields in the solar chromosphere. *Nature* **392**, 256–258 (1998).
25. Stenflo, J. O. in *Solar Polarization 3* (eds Trujillo Bueno, J. & Sánchez Almeida, J.) 385–398 (ASP Conf. Ser. Vol. 307, Astronomical Society of the Pacific, San Francisco, 2003).
26. Anderson, L. S. & Athay, R. G. Model solar chromosphere with prescribed heating. *Astrophys. J.* **346**, 1010–1018 (1989).
27. Priest, E. & Forbes, T. *Magnetic Reconnection: MHD Theory and Applications* (Cambridge Univ. Press, New York, 2000).
28. Landi Degl'Innocenti, E. Polarization in spectral lines: I. A unifying theoretical approach. *Sol. Phys.* **85**, 3–31 (1983).
29. Trujillo Bueno, J. in *Stellar Atmosphere Modeling* (eds Hubeny, I., Mihalas, D. & Werner, K.) 551–582 (ASP Conf. Ser. Vol. 288, Astronomical Society of the Pacific, San Francisco, 2003).
30. Trujillo Bueno, J. in *Solar Polarization 3* (eds Trujillo Bueno, J. & Sánchez Almeida, J.) 407–424 (ASP Conf. Ser. Vol. 307, Astronomical Society of the Pacific, San Francisco, 2003).

Acknowledgements We thank F. Kneer, E. Landi Degl'Innocenti and F. Moreno-Insertis for scientific discussions. We are also grateful to P. Fabiani Bendicho for help with the numerical solution of the 3D radiative transfer equation. This research was supported by the Spanish Plan Nacional de Astronomía y Astrofísica and by the European Commission via the INTAS programme and the Solar Magnetism Network.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T.B. (jtb@iac.es).

Single spin detection by magnetic resonance force microscopy

D. Rugar, R. Budakian, H. J. Mamin & B. W. Chui

IBM Research Division, Almaden Research Center, 650 Harry Rd, San Jose, California 95120, USA

Magnetic resonance imaging (MRI) is well known as a powerful technique for visualizing subsurface structures with three-dimensional spatial resolution. Pushing the resolution below 1 μm remains a major challenge, however, owing to the sensitivity limitations of conventional inductive detection techniques. Currently, the smallest volume elements in an image must contain at least 10^{12} nuclear spins for MRI-based microscopy¹, or 10^7

electron spins for electron spin resonance microscopy². Magnetic resonance force microscopy (MRFM) was proposed as a means to improve detection sensitivity to the single-spin level, and thus enable three-dimensional imaging of macromolecules (for example, proteins) with atomic resolution^{3,4}. MRFM has also been proposed as a qubit readout device for spin-based quantum computers^{5,6}. Here we report the detection of an individual electron spin by MRFM. A spatial resolution of 25 nm in one dimension was obtained for an unpaired spin in silicon dioxide. The measured signal is consistent with a model in which the spin is aligned parallel or anti-parallel to the effective field, with a rotating-frame relaxation time of 760 ms. The long relaxation time suggests that the state of an individual spin can be monitored for extended periods of time, even while subjected to a complex set of manipulations that are part of the MRFM measurement protocol.

MRFM is based on the detection of the magnetic force between a ferromagnetic tip and spins in a sample. The fundamental challenge in achieving single-spin sensitivity is that the force from a single spin is exceedingly small. Even with tip field gradients in the gauss per nanometre range, the force from an electron spin is only a few attonewtons. This force is roughly a million times smaller than is typically detected by atomic force microscopy (AFM). Recently, major strides towards single-spin detection have been made with the development of ultrasensitive cantilever-based force sensors^{7,8}, better understanding of relevant spin relaxation processes^{9–12} and the successful detection of statistical polarization in small spin ensembles¹³.

The basic elements of our MRFM apparatus are shown in Fig. 1 and have been described in detail in ref. 13. Briefly, a custom-fabricated mass-loaded silicon cantilever^{8,13} with an attached 150-nm-wide SmCo magnetic tip is used to sense the force from the electron spin. The sample consists of vitreous silica (Suprasil W2) that was irradiated with a 2-Gy dose of Co^{60} gamma rays. The gamma irradiation produces a low concentration of silicon dangling bonds containing unpaired electron spins known as E' centres¹⁴.

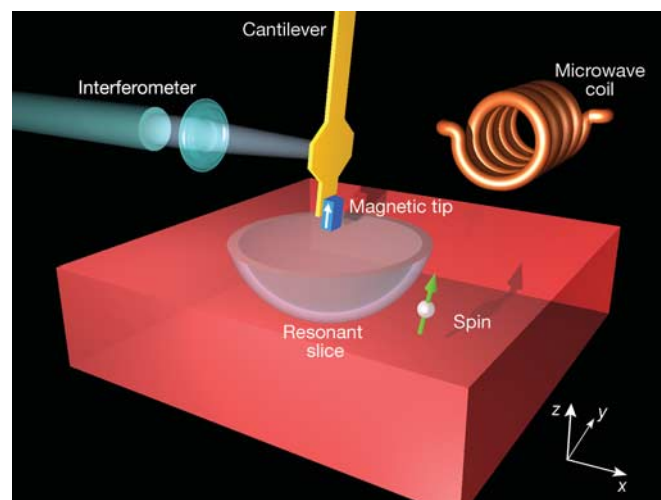


Figure 1 Configuration of the single-spin MRFM experiment. The magnetic tip at the end of an ultrasensitive silicon cantilever is positioned approximately 125 nm above a polished SiO_2 sample containing a low density of unpaired electron spins. The resonant slice represents those points in the sample where the field from the magnetic tip (plus an external field) matches the condition for magnetic resonance. As the cantilever vibrates, the resonant slice swings back and forth through the sample causing cyclic adiabatic inversion of the spin. The cyclic spin inversion causes a slight shift of the cantilever frequency owing to the magnetic force exerted by the spin on the tip. Spins as deep as 100 nm below the sample surface can be probed.

Estimated spin concentration was between 10^{13} and 10^{14} cm^{-3} . The experiments were performed at 1.6 K in a small vacuum chamber that fits within the bore of a superconducting magnet. The low operating temperature minimizes the force noise and reduces the relaxation rate of the spins.

A microwave magnetic field ($B_1 \approx 0.3 \text{ mT}$), in combination with the inhomogeneous field from the magnetic tip, sets up a 'resonant slice' within the sample. The resonant slice corresponds to those points in the sample where the tip field $\mathbf{B}_{\text{tip}}(x, y, z)$, plus a static external field $\hat{z} B_{\text{ext}}$, satisfies the condition for electron spin resonance: $B_0(x, y, z) \equiv |\mathbf{B}_{\text{tip}}(x, y, z) + \hat{z} B_{\text{ext}}| = \omega_{\text{rf}}/\gamma$. Here, ω_{rf} is the frequency of the microwave field and γ is the gyromagnetic ratio ($\gamma/2\pi = 2.8 \times 10^{10} \text{ Hz T}^{-1}$). In the present experiment, where $\omega_{\text{rf}}/2\pi = 2.96 \text{ GHz}$, the resonant slice corresponds to $B_0(x, y, z) = 106 \text{ mT}$. For typical conditions (for example, $B_{\text{ext}} = 30 \text{ mT}$), the slice is a bowl-shaped surface that extends roughly 250 nm below the tip. Note that with the perpendicular cantilever orientation shown in Fig. 1, the cantilever is responsive only to the x component of force. Thus the spin must be located either slightly in front of or slightly behind the cantilever in the x direction to produce a substantial response.

To generate a force signal that can be distinguished from the much larger background force fluctuations, we use the recently developed spin manipulation protocol known as 'interrupted OSCAR' or iOSCAR, where OSCAR stands for oscillating cantilever-driven adiabatic reversals¹³. In the iOSCAR protocol, the cantilever is part of a gain-controlled positive-feedback loop that drives the cantilever to oscillate at a set amplitude (for example, 16 nm) at the fundamental frequency of the cantilever ($f_c = 5.5 \text{ kHz}$). Because the cantilever is the frequency-determining element in the feedback loop, the vibration frequency will automatically vary in response to tip-sample interactions¹⁵.

The vibration of the cantilever tip causes the resonant slice to sweep back and forth rapidly through the sample. If the slice happens to sweep through the location of a spin, the spin will be cyclically inverted in synchrony with the cantilever motion owing to the phenomenon of adiabatic rapid passage^{13,16}. This synchronous inversion of the spin creates an alternating magnetic force on the cantilever that mimics a change in cantilever stiffness. The resulting shift in cantilever frequency is given by^{13,17}:

$$\delta f_c = \pm \frac{2f_c G \mu_B}{\pi k x_{\text{peak}}}, \quad (1)$$

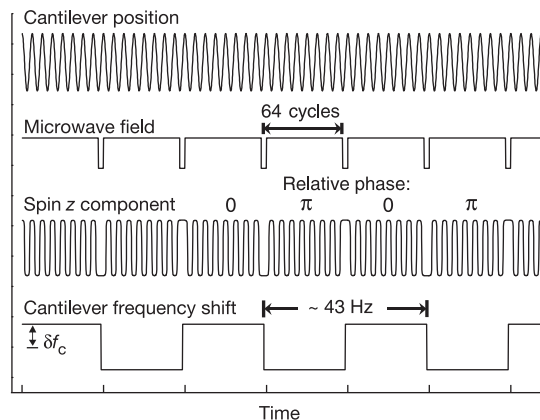


Figure 2 Timing diagram for the iOSCAR spin manipulation protocol. The z component of spin follows the motion of the cantilever except during the interruption of the microwave field. The interruptions last for one-half of a cantilever period and are precisely timed to start and stop at vibration extrema. The interruptions, which occur every 64 cantilever cycles, reverse the phase of the spin with respect to the cantilever, resulting in a square-wave oscillation of the cantilever frequency.

where k is the cantilever spring constant, x_{peak} is the peak vibration amplitude of the cantilever, $G \equiv \partial B_0/\partial x$ is the lateral field gradient, and μ_B is the magnetic moment of the electron ($9.3 \times 10^{-24} \text{ J T}^{-1}$). The sign of the frequency shift depends on the relative phase of the spin inversions with respect to the cantilever motion. Using the terminology of magnetic resonance, the two polarities correspond to adiabatic rapid passages with the spin either aligned or anti-aligned with respect to the effective field in the rotating frame^{13,16}. This parallel or anti-parallel alignment is expected to be enforced by the quantum-mechanical measurement process, resulting in the collapse of the initial spin wavefunction onto an eigenstate of the effective field^{18,19}. For the parameters of the current experiment ($G = 2 \times 10^5 \text{ T m}^{-1}$; $k = 0.11 \text{ mN m}^{-1}$; $x_{\text{peak}} = 16 \text{ nm}$), the expected frequency shift is $|\delta f_c| = 3.7 \pm 1.3 \text{ mHz}$. The estimated uncertainty in $|\delta f_c|$ reflects 20% uncertainties in the calibration of G , k and x_{peak} .

In the 'interrupted' version of the protocol (iOSCAR), the microwave field B_1 is turned off for one-half of a cantilever cycle every 64 cycles ($f_{\text{int}} = f_c/64 \approx 86 \text{ Hz}$). As shown in Fig. 2, each time B_1 is interrupted, the relative phase of the spin and cantilever is reversed, causing the frequency shift to reverse polarity. (See Supplementary Information for an animation.) The net result of the interruptions is that the frequency shift will alternate between positive and negative values in a square-wave-like fashion with a frequency given by $f_{\text{sig}} = f_{\text{int}}/2$, or approximately 43 Hz. The resulting frequency shift signal can be written as the Fourier series:

$$\Delta f(t) = \frac{4}{\pi} |\delta f_c| A(t) \sin(2\pi f_{\text{sig}} t) + \text{higher harmonics} \quad (2)$$

where the $4/\pi$ comes from the first harmonic Fourier amplitude of a square wave.

We have included the function $A(t)$ to take into account the fact that the signal will not be perfectly periodic owing to extra random spin flips induced by the environment. Assuming that this 'quantum jump' model is correct, $A(t)$ is a random telegraph function that takes on values of ± 1 . For Poisson distributed jumps, $A(t)$ is expected to have a lorentzian power spectrum²⁰ and the statistical properties $\langle A(t) \rangle = 0$ and $\langle [A(t)]^2 \rangle = 1$, where $\langle \dots \rangle$ denotes time average. We detect the first harmonic of the signal only, so we will

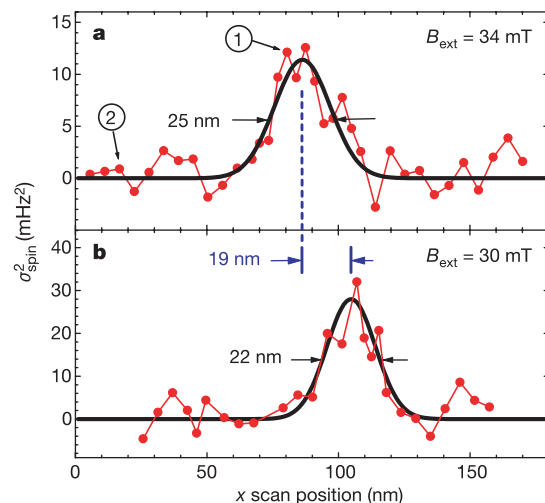


Figure 3 Plots showing the spin signal as the sample was scanned laterally in the x direction for two values of external field: **a**, $B_{\text{ext}} = 34 \text{ mT}$, and **b**, $B_{\text{ext}} = 30 \text{ mT}$. The smooth curves are gaussian fits that serve as guides to the eye. The 19-nm shift in peak position reflects the movement of the resonant slice induced by the 4-mT change in external field. The difference in absolute peak height is primarily due to different lock-in amplifier detection bandwidths: 0.18 Hz and 0.59 Hz for **a** and **b**, respectively. Power spectra for the points marked 1 and 2 are shown in Fig. 4.

refer to the quantity $\Delta f_1(t) \equiv (4/\pi)|\delta f_c|A(t)$ as the spin signal amplitude.

Because the frequency modulation due to the spin is only a few millihertz and the frequency noise of the cantilever due to thermal motion and tip-sample interactions is relatively large (~ 25 mHz in a 1-Hz bandwidth), we must use signal averaging to detect the spin signal. However, because $\langle \Delta f_1(t) \rangle = 0$, we average the square of the signal (the signal 'energy'), rather than the signal amplitude. Specifically, the frequency modulation of the cantilever is detected using an analogue frequency discriminator¹⁵ followed by a digital lock-in amplifier that has been implemented in software. The lock-in amplifier uses a bank of low-pass filters and, as a function of detection bandwidth, determines the energy (that is, variance) of the in-phase and quadrature components of the frequency-shift signal $\Delta f(t)$. The spin signal and the measurement noise are uncorrelated, so we can write the in-phase variance as $\sigma_I^2 = \sigma_{\text{spin}}^2 + \sigma_{\text{noise}}^2$, where σ_{spin}^2 is the variance due to the spin signal and σ_{noise}^2 represents the measurement noise. This in-phase variance is then compared to the quadrature variance σ_Q^2 , which contains only the measurement noise. The signal energy from the spin can then be estimated as $\sigma_{\text{spin}}^2 = \sigma_I^2 - \sigma_Q^2$. This energy detection scheme

is believed to have performance close to the theoretically optimal detector²¹.

Figure 3a shows a lateral scan where we plot σ_{spin}^2 as a function of sample position. The scan shows a prominent peak that we believe is due to a single spin. The peak width is 25 nm, or roughly 1.6 times the cantilever oscillation amplitude. Note that the baseline of the data on either side of the peak is essentially zero (within the uncertainty of the measurement), thus demonstrating that the iOSCAR protocol generates no systematic feedthrough artefacts. Because the signal-to-noise ratio was so low ($\sigma_{\text{spin}}^2/\sigma_I^2 \sim 0.06$), considerable averaging was required. In Fig. 3a, the averaging time was 13 h per point, yielding a signal peak that is five standard deviations above the baseline noise.

To confirm that the observed signal is truly due to magnetic resonance, a number of basic tests were performed. As expected for an iOSCAR magnetic resonance signal, the signal disappeared if the microwaves were absent or turned on continuously. The timing of the microwave interruptions was also varied. The signal disappeared, as expected, when the starting time of the interruption was shifted from the vibration peak to the zero-crossing of the vibration (that is, shifted by one-quarter of a cantilever cycle) and when the interruption duration was a full cantilever cycle instead of one-half cycle.

One key test for magnetic resonance is to observe the field dependence of the spin signal. If the external field is reduced, the resonant slice will shrink in radius, thus shifting the scan position of the signal peak. When B_{ext} was reduced from 34 to 30 mT, the expected peak shift was indeed observed and found to be $\Delta x = 19$ nm, as can be seen by comparing the scans in Figs 3a and b. The ratio $\Delta B_{\text{ext}}/\Delta x$ suggests that the field gradient G is approximately $2 \times 10^5 \text{ T m}^{-1}$ (2 G nm^{-1}).

The conclusion that the signal is due to only one spin is based primarily on the spatial isolation of the spin signal. By design, the spin density of the sample was very low, in the range of 10^{13} to 10^{14} cm^{-3} , giving a mean spacing between spins in the range of 200 to 500 nm. The sparseness of spins implies that, for most sample locations, there is no spin interacting with the resonant slice. This is why the data in Fig. 3 has a zero baseline. To locate a spin signal, the sample was scanned through many independent locations, of the order of 30, before a strong signal from a well-positioned spin was found.

By measuring σ_{spin}^2 as a function of detection bandwidth, the power spectral density of the spin signal amplitude $\Delta f_1(t)$ can be determined (Fig. 4a). Consistent with a random telegraph model of $\Delta f_1(t)$, the spectrum is well fitted by the lorentzian function $S(f) = 4\tau_m \langle [\Delta f_1(t)]^2 \rangle / [1 + 4\pi^2 \tau_m^2 f^2]$. The spectral width at half-maximum is 0.21 Hz, corresponding to an impressively long τ_m correlation time of 760 ms. This is essentially the rotating frame relaxation time, but also includes possible effects of the iOSCAR spin manipulation. The long correlation time implies that the cantilever-driven spin inversions are coherent for thousands of cycles. The false-colour image in Fig. 4b shows that the spin signal is highly localized both spatially and spectrally.

The total magnitude of the spin signal, obtained by integrating the spectrum in Fig. 4a, was found to be $\langle [\Delta f_1(t)]^2 \rangle = 28 \text{ mHz}^2$. Using this experimental value and the relationship $\langle [\Delta f_1(t)]^2 \rangle = (4/\pi)^2 |\delta f_c|^2 \langle [A(t)]^2 \rangle$, we solve for $|\delta f_c|$ using the assumption that $\langle [A(t)]^2 \rangle = 1$. We find $|\delta f_c| = 4.2 \text{ mHz}$, in excellent agreement with the value of 3.7 mHz expected from equation (1).

In conclusion, we have presented evidence that MRFM is now capable of detecting individual electron spins. Although several other single-spin detection methods have been previously demonstrated^{22–28}, MRFM has some attributes that set it apart. Perhaps the most important of these is the ability to image spins below the surface with nanometre spatial resolution. Spins as deep as 100 nm should be accessible under present operating conditions. Although extensive signal averaging is currently required, even a modest

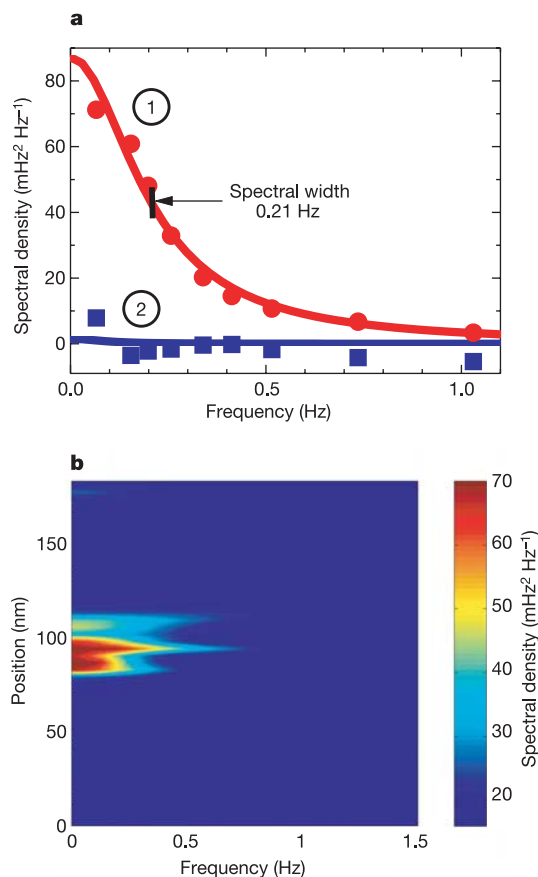


Figure 4 By measuring the spin signal energy as a function of detection bandwidth, the power spectral density of the spin signal amplitude $\Delta f_1(t)$ can be determined. **a**, Spectral density for the two positions indicated in Fig. 3. The strong spin signal at position 1 has a narrow spectral width (0.21 Hz), reflecting the long (760 ms) correlation time obtained using iOSCAR spin manipulation. The data are well fitted by a Lorentzian function (solid curve). At position 2, approximately 70 nm away from the position of the spin signal, the spectral density is negligible. **b**, False-colour plot showing power spectral density as a function of position. The spin signal is localized both spatially and spectrally. The data were interpolated between discrete measurement positions to create a smooth image. The apparent fine structure in the signal region is probably an artefact of the limited signal-to-noise ratio of the original data.

increase in field gradient (for example, five times larger) will dramatically speed up the acquisition time and thereby enable two- and three-dimensional imaging applications. (Because we average signal energy, rather than amplitude, the averaging time will decrease as the inverse fourth power of the gradient in the limit of low signal-to-noise ratio.) If the measurement time can be reduced below τ_m , real-time readout of the spin quantum state will become possible, enabling a wide variety of quantum measurement experiments. For molecular imaging applications, extension to single nuclear spin detection is necessary, but this will require a roughly 1,000-fold improvement in magnetic moment sensitivity. Considering that the present experiment represents a sensitivity improvement of 10^7 times over the original MRFM experiment²⁹, the remaining required improvement does not seem out of the question, especially since there is still considerable leeway for increasing the field gradient and lowering the operating temperature. □

Received 16 April; accepted 17 May 2004; doi:10.1038/nature02658.

1. Ciobanu, L., Seeber, D. A. & Pennington, C. H. 3D MR microscopy with resolution $3.7 \mu\text{m}$ by $3.3 \mu\text{m}$. *J. Magn. Reson.* **158**, 178–182 (2002).
2. Blank, A., Dunnam, C. R., Borbat, P. P. & Freed, J. H. High resolution electron spin resonance microscopy. *J. Magn. Reson.* **165**, 116–127 (2003).
3. Sidles, J. A. Folded Stern-Gerlach experiment as a means for detecting nuclear magnetic resonance in individual nuclei. *Phys. Rev. Lett.* **68**, 1124–1127 (1992).
4. Sidles, J. A. *et al.* Magnetic resonance force microscopy. *Rev. Mod. Phys.* **67**, 249–265 (1995).
5. DiVincenzo, D. P. Two-bit gates are universal for quantum computation. *Phys. Rev. A*, **51**, 1015–1022 (1995).
6. Berman, G. P., Doolen, G. D., Hammel, P. C. & Tsifrinovich, V. I. Solid-state nuclear-spin quantum computer based on magnetic resonance force microscopy. *Phys. Rev. B* **61**, 14694–14699 (2000).
7. Stowe, T. D. *et al.* Attonewton force detection using ultrathin silicon cantilevers. *Appl. Phys. Lett.* **71**, 288–290 (1997).
8. Chui, B. W. *et al.* Mass-loaded cantilevers with suppressed higher-order modes for magnetic resonance force microscopy. *Technical Digest 12th Int. Conf. on Solid-State Sensors and Actuators (Transducers'03)* 1120–1123 (IEEE, Piscataway, 2003).
9. Stipe, B. C. *et al.* Electron spin relaxation near a micron-size ferromagnet. *Phys. Rev. Lett.* **87**, 277602 (2001).
10. Mozysky, D., Martin, I., Pelekhov, D. & Hammel, P. C. Theory of spin relaxation in magnetic resonance force microscopy. *Appl. Phys. Lett.* **82**, 1278–1280 (2003).
11. Berman, G. P., Gorshkov, V. N., Rugar, D. & Tsifrinovich, V. I. Spin relaxation caused by thermal excitations of high-frequency modes of cantilever vibration. *Phys. Rev. B* **68**, 094402 (2003).
12. Hannay, J. D., Chantrell, R. W. & Rugar, D. Thermal field fluctuations in a magnetic tip—implications for magnetic resonance force microscopy. *J. Appl. Phys.* **87**, 6827–6829 (2000).
13. Mamin, H. J., Budakian, R., Chui, B. W. & Rugar, D. Detection and manipulation of statistical polarization in small spin ensembles. *Phys. Rev. Lett.* **91**, 207604 (2003).
14. Castle, J. G., Feldman, D. W., Klemens, P. G. & Weeks, R. A. Electron spin-lattice relaxation at defect sites: E' centers in synthetic quartz at 3 kilo-Oersteds. *Phys. Rev.* **130**, 577–588 (1963).
15. Albrecht, T. R., Grütter, P., Horne, D. & Rugar, D. Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **69**, 668–673 (1991).
16. Slichter, C. P. *Principles of Magnetic Resonance*, 3rd edn, 20–24 (Springer, Berlin, 1990).
17. Berman, G. P., Kamenov, D. I. & Tsifrinovich, V. I. Stationary cantilever vibrations in oscillating-cantilever-driven adiabatic reversals: Magnetic-resonance-force-microscopy technique. *Phys. Rev. A*, **66**, 023405 (2002).
18. Berman, G. P., Borgonovi, F., Goan, H.-S., Gurvitz, S. A. & Tsifrinovich, V. I. Single-spin measurement and decoherence in magnetic-resonance force microscopy. *Phys. Rev. B* **67**, 094425 (2003).
19. Brun, T. A. & Goan, H.-S. Realistic simulations of single-spin nondemolition measurement by magnetic resonance force microscopy. *Phys. Rev. A*, **68**, 032301 (2003).
20. Davenport, W. B. & Root, W. L. *An Introduction to the Theory of Random Signals and Noise* 104 (McGraw-Hill, New York, 1958).
21. Ting, M., Hero, A. O., Rugar, D., Yip, C.-Y. & Fessler, J. A. Electron spin detection in the frequency domain under the interrupted oscillating cantilever-driven adiabatic reversal (iOSCAR) protocol. Preprint at <http://xxx.lanl.gov/abs/quant-ph/0312139> (2003).
22. Manassen, Y., Hamers, R. J., Demuth, J. E. & Castellano, A. J. Direct observation of the precession of individual paramagnetic spins on oxidized silicon surfaces. *Phys. Rev. Lett.* **62**, 2531–2534 (1989).
23. Durkan, C. & Welland, M. E. Electronic spin detection in molecules using scanning-tunneling-microscopy-assisted electron-spin resonance. *Appl. Phys. Lett.* **80**, 458–460 (2002).
24. Wrachtrup, J., von Borczyskowski, C., Bernard, J., Orritt, M. & Brown, R. Optical-detection of magnetic resonance in a single molecule. *Nature* **363**, 244–245 (1993).
25. Köhler, J. *et al.* Magnetic resonance of a single molecular spin. *Nature* **363**, 242–244 (1993).
26. Jelezko, F. *et al.* Single spin states in a defect center resolved by optical spectroscopy. *Appl. Phys. Lett.* **81**, 2160–2162 (2002).
27. Elzerman, J. M. *et al.* Single shot read-out of an individual electron spin in a quantum dot. *Nature* (in the press).
28. Jiang, H.-W., Xiao, M., Martin, I. & Yablonovitch, E. Electrical detection of electron spin resonance of a single spin in the SiO₂ of a Si field effect transistor. *Nature* (in the press).
29. Rugar, D., Yannoni, C. S. & Sidles, J. A. Mechanical detection of magnetic resonance. *Nature* **360**, 563–566 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Sidles, A. Hero, M. Ting, G. Berman, I. Martin, C. S. Yannoni and T. Kenny for discussions, and D. Pearson, Y. Hishinuma, M. Sherwood and C. Rettner for technical assistance. This work was supported by the DARPA Three-Dimensional Atomic-Scale Imaging programme administered through the US Army Research Office.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R. (rugar@almaden.ibm.com).

Formation of zirconium metallic glass

Jianzhong Zhang & Yusheng Zhao

LANSCE Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Bulk metallic glasses are commonly produced by the rapid cooling of liquid alloys¹. They have emerged over the past decade as a novel class of materials, with attractive properties and technological promise^{1,2}. The bulk metallic glasses so far produced contain three or more component elements^{3,4}. These complex compositions are necessary to frustrate the crystallization of the liquid melt on cooling, but can also lead to phase separation, which is detrimental to the thermal and mechanical properties of metallic glasses^{5–8}. Here we report, using X-ray diffraction measurements, the formation of a bulk metallic glass from elemental zirconium at high static pressures and low temperatures (relative to its melting temperature at atmospheric pressure). Amorphous zirconium can be recovered at ambient conditions and demonstrates a superior thermal stability compared to amorphous alloys^{3,9}, which could lead to new high-temperature applications of amorphous metals.

In multi-component systems, glass-forming ability (GFA) is viewed as the resistance to precipitation of crystalline phases from supercooled liquid metals¹⁰, and alloys with high GFA all have three common features^{3,4}: (1) they consist of at least three components; (2) there is significant mismatch of the atomic size of the constituent elements; and (3) there are negative heats of mixing among the major alloying elements. Addition of elements that are chemically and topologically different from the other species not only creates an energy barrier for nuclei to form but also effectively increases melt viscosity or fragility, which results in a reduced rate of both nucleation and growth and an increase in GFA. The production of bulk glassy materials in pure metals, however, remains a long-standing scientific curiosity and technological interest. The difficulties arise from the facts that the equilibrium melt viscosity of pure metals is three orders of magnitude smaller than that of amorphous alloys¹¹ and that current technology has yet to reach a cooling rate in excess of the $10^{10} \text{ °C s}^{-1}$ that is needed to make pure metals amorphous¹².

We studied zirconium metal at pressures and temperatures up to 17 GPa and 1,000 °C, using energy-dispersive synchrotron X-ray diffraction and time-of-flight neutron scattering. In X-ray diffraction experiments, we used both DIA-type¹³ and T-cup¹⁴ large-volume high-pressure apparatus installed at Brookhaven and Argonne National Laboratories. Neutron scattering experiments were performed using a high-pressure/high-temperature (high *P*–*T*) cell assembly¹⁵ in a TAP-98 toroidal-anvil press at Los Alamos Neutron Scattering Center. The starting sample of zirconium has a close-packed hexagonal structure (α -phase) and is of extremely high purity, with 35 p.p.m. Hf, less than 25 p.p.m. of C, N and Al,

and less than 50 p.p.m. of O, V and Fe. The zirconium samples were placed in boron nitride and/or NaCl capsules, with NaCl used as an internal pressure standard¹⁶.

Both X-ray and neutron diffraction experiments reveal that the zirconium metal is in crystalline forms at pressures and temperatures up to 4.3 GPa and 900 °C. The formation of amorphous zirconium was first observed when ω -phase (a hexagonal structure but not close-packed) was heated to 650 °C at 5.3 GPa (Figs 1 and 2). The lower bound for glass formation is therefore bracketed between pressures of 4.3 and 5.3 GPa. Upon cooling, we observed crystallization of ω -phase from amorphous zirconium at 4.8 GPa and 450 °C, indicating that the crystalline–glass transformation is reversible at this P – T condition (Fig. 2). When zirconium metal was heated at higher pressures of 6.4 GPa and 8.6 GPa in two different experiments, glass formation was observed at 700 °C and 625 °C, respectively, and the glass was stable on further heating to 1,000 °C (Figs 1 and 3). In both experiments, the transformation sequence is from ω -phase to β -phase (a body-centred cubic, high-temperature phase), and from β -phase to amorphous phase. The temperature interval between first appearance of β and first appearance of amorphous phase is less than 25 °C. Upon slow cooling, we did not observe crystallization of either ω - or β -phase even at room temperature, indicating an irreversible process for glass formation in zirconium metal. These findings demonstrate that pressure plays a critical role in both formation and reversibility of amorphous zirconium. From another perspective, we can conclude that the irreversible formation of amorphous zirconium seems to occur only when glass is formed from the β -phase.

To further explore effect of pressure on the GFA, we performed an additional experiment at 14–17 GPa. We observed the transformation from ω -phase to β -phase between 500 and 600 °C; upon

further heating, however, zirconium metal is only partially transformed to an amorphous phase at 750 °C and 14.5 GPa, and the diffraction pattern is dominated by the β -phase. Compared with our experimental results at lower pressures (Fig. 1), these observations tend to suggest that zirconium metal has a diminishing GFA with increasing pressure between 9 and 15 GPa. The fully amorphous zirconium can therefore be formed in a limited pressure–temperature space. The glass-forming region needs to be more accurately determined to constrain the phase diagram for zirconium metal, as well as to provide a better understanding of glass-forming kinetics.

The glass formed at 8.6 GPa and temperatures above 625 °C (Fig. 3) was annealed up to 900 °C at pressures as low as 2.8 GPa to investigate the thermal stability of amorphous zirconium outside its formation P – T conditions (Fig. 4). At these experimental conditions, we did not observe the precipitation of any crystalline phase from amorphous zirconium. The glass formed at high pressure and temperature appears to have a superior thermal stability when compared to amorphous alloys formed from the conventional melting–cooling process, which typically start crystallization when annealed at temperatures between 450 and 500 °C (refs 3, 9). Whether amorphous zirconium shows a similar thermal behaviour at atmospheric pressure remains to be explored. As a final note on the stability, both energy and angle dispersive diffraction on the recovered zirconium sample shows patterns that are characteristic of an amorphous phase, but they do not reveal any diffraction lines of crystalline zirconium (Figs 2–4 and Supplementary Fig. 1). Both recovered and starting samples do not show any Raman shifts except for surface oxidation, indicating that the bulk zirconium sample remains in an elemental form at our experimental con-

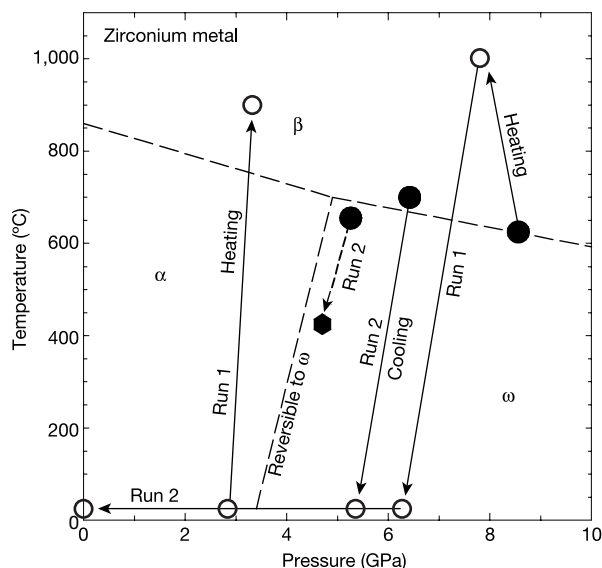


Figure 1 Glass-forming conditions and stability of amorphous zirconium. Formation of glass was observed in two independent experiments (runs 1 and 2). Filled circles refer to pressures and temperatures at which crystalline zirconium metal fully transformed into an amorphous phase on heating. The filled hexagon symbol corresponds to the conditions where the reversed transformation from glass to ω -phase was observed upon cooling. Arrows indicate the selected experimental P – T paths along which no precipitation of any crystalline phase was observed after the formation of amorphous zirconium, with open circles denoting a portion of data collected along the paths (see Fig. 4b for more detail for run 1). Experimental constraints on the phase diagram of zirconium metal (dashed lines) will be published elsewhere. A small amount of elemental addition has previously been demonstrated to significantly affect the glass-forming ability and physical properties in multi-component alloys⁹. Such effects in zirconium metal need to be explored as this study was carried out on zirconium specimens of ultra-high purity.

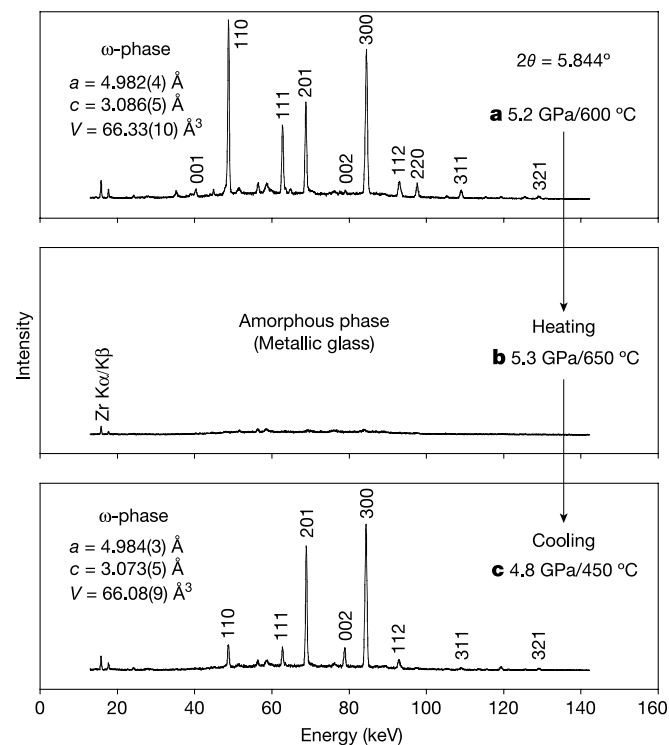


Figure 2 Selected synchrotron X-ray diffraction patterns showing a reversible transformation between ω -phase and amorphous phases of zirconium metal. Formation of a glass is identified by the disappearance of all diffraction peaks of ω -phase and by a significant reduction in the diffraction intensity (pattern b). Intensities have been normalized to the data acquisition time, and all three figures are plotted within an identical intensity range. The weak diffraction peaks observed in the amorphous phase are also present in all patterns collected in this experiment, and are from diffraction of materials surrounding the zirconium metal (see patterns a and c).

ditions. The microstructure and/or atomic scale structure of the recovered sample, however, need to be characterized by other experimental techniques, such as high-resolution transmission electron microscopy and neutron pair distribution function.

Phase transformations in zirconium metal have previously been studied at pressures and temperatures up to 6 GPa and 840 °C (ref. 17), and at 31–36 GPa up to 175 °C (refs 18, 19). The experimental conditions we found for the irreversible glass formation have therefore not been explored by any previous work. In addition, no theoretical calculations have predicted the instability of ω - and β -phases of zirconium metal relative to a solid amorphous phase^{20,21}. At atmospheric pressure, zirconium metal melts at 1,855 °C; the melting curve at high pressures, however, is not known. But it is extremely unlikely that the observed glass formation can be attributed to the melting of zirconium metal at the experimental pressures, because this would require an unusually large negative dT/dP slope of ~ 240 °C GPa⁻¹. The melting explanation is made more unlikely by the fact that the observed glass-forming temperature between 5 and 9 GPa only has a slightly

negative dependence on pressure (Fig. 1). We therefore conclude that glass formation in zirconium metal by-passes the conventionally required liquid state, which represents a novel approach to making metallic glass. This experimental study presents (to our knowledge) the first finding that bulk metallic glass can be produced from a pure elemental metal.

Formation of an amorphous phase within a solid state by application of pressure has been found to be a phenomenon of widespread occurrence among condensed matter systems, such as silicates²², ice²³ and alloys²⁴. There are, however, major differences between these observations and the present findings. Pressure-induced amorphization is a process in which amorphous phase is formed because transformation from a crystalline solid to its high-pressure phase is kinetically hindered at ambient or low temperature²⁵. In zirconium metal, formation of a glass occurs after α -phase transforms to its high-pressure and/or high-temperature phases. For pressure-induced amorphization, the glass would crystallize into its equilibrium, high-pressure phase when subjected to elevated temperatures. In zirconium glass, no precipitation of any crystalline phases was found at temperatures up to 1,000 °C. In fact, it is not known that there exists another high-temperature polymorph of zirconium metal except the β -phase. This leads to an unusual situation where amorphous zirconium would eventually 'melt' at somewhat higher temperatures, which could infer an extremely high thermal stability. Finally, no previous work has demonstrated that pure metals would undergo a pressure-induced amorphization.

Both measurements and molecular dynamics have revealed strong or anomalous softening of phonons along high symmetry directions in crystals of α -, ω - and β -phases of zirconium metal^{126–28}. In particular, phonon anomalies have been observed in the dis-

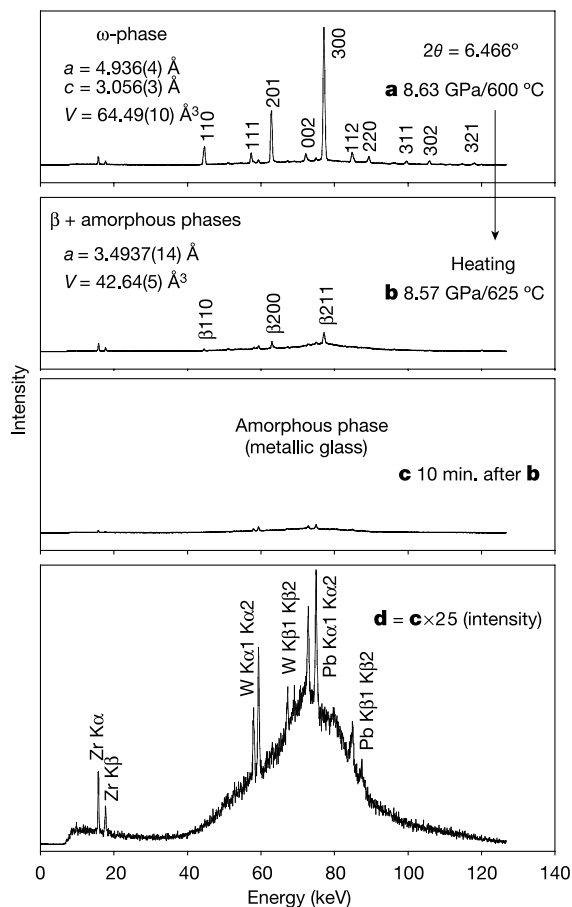


Figure 3 Selected diffraction patterns showing the ω - β phase transition and formation of glass from the β -phase of zirconium metal at 8.6 GPa. Pattern **d** is a blow-up of pattern **c**. It is well known that β -phase is formed by the splitting of alternating (001) planes along the c axis of an ω structure into two (111) planes of the β -phase. Therefore, diffraction pattern of the ω -phase contains all the diffraction lines of the β -phase and some characteristic lines resulting from its superlattice structure (patterns **b** and **c**). In this study and previous ones^{18,19}, the transition from ω -phase to β -phase is identified by the disappearance of the superlattice diffraction lines, (111), (002) and (112), of the ω -phase. Weak diffraction lines observed in all patterns are fluorescence peaks of tungsten, which was used to define collimation of the multi-element detector, and lead, which was used as a shielding material to avoid air scattering. For patterns **a**, **b** and **c**, intensities have been normalized to the data acquisition time, and all three figures are plotted within an identical intensity range.

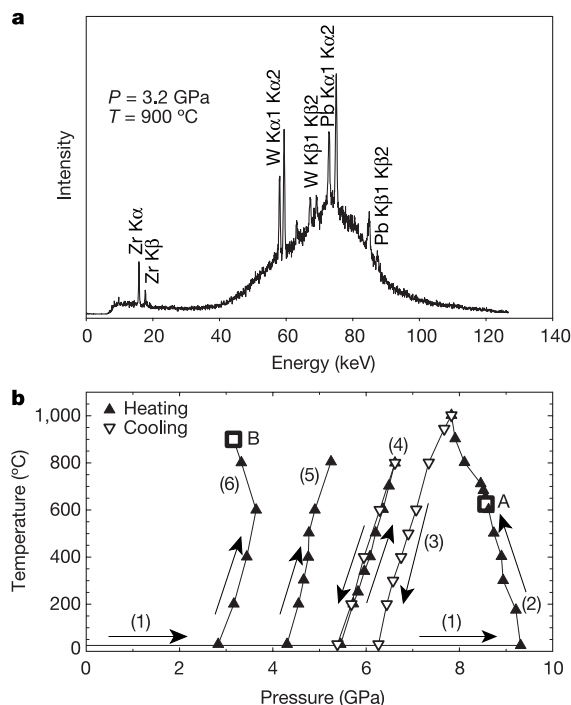


Figure 4 Experimental results demonstrating thermal stability/metastability of zirconium metallic glass. **a**, Diffraction pattern of glassy zirconium collected outside its forming conditions (Fig. 1). **b**, The P - T path for the experimental run 1. Open squares correspond to the conditions at which diffraction data were taken for Fig. 3d (point A) and Fig. 4a (point B). Numbers in parentheses and arrows indicate the sequential development of data collection in this experiment. After the first formation of amorphous zirconium at point A (8.57 GPa, 625 °C), the patterns collected in subsequent heating-cooling cycles at lower pressures and at the end of experiment are essentially identical to the ones shown in Fig. 3d and Fig. 4a.

persion curve of the high-temperature β -phase of zirconium at ambient pressure, which has been related to the $\beta \rightarrow \omega$ and $\beta \rightarrow \alpha$ phase transitions²⁶. It is therefore possible that these phonon mode softenings may play a key role in the formation of zirconium metallic glass. Inelastic scattering at elevated pressure and temperature is needed to investigate the driving mechanisms for the present observations.

Liquid–liquid phase separation or decomposition of deeply undercooled metallic liquids is at present a major problem faced in the fundamental study and technological processing of bulk metallic glasses, even for extremely good glass-formers like Vit1 ($\text{Zr}_{41.2}\text{Ti}_{13.8}\text{Cu}_{12.5}\text{Ni}_{10}\text{Be}_{22.5}$) and Vit105 ($\text{Zr}_{52.5}\text{Ti}_{5}\text{Cu}_{17.9}\text{Ni}_{14.6}\text{Al}_{10}$) (refs 5–8). Similarly, decomposition happens when amorphous alloys are annealed at temperatures above the glass transition temperature⁵, T_g . Experiments have revealed that such decomposition is responsible for the embrittlement of some bulk metallic glasses^{1,5}. For mechanical applications, it is important to find metallic glasses that have higher thermal stability (that is, a smaller difference between the critical temperature¹, T_c , and T_g) and thus a lower probability of decomposition. As $T_c - T_g$ reduces, however, the amorphous alloys, such as derivatives of Vit1 and Vit105, tend to have a somewhat diminished GFA. We consider that the extraordinary GFA of pure zirconium metal, combined with its superior thermal stability, will overcome the problems that exist in amorphous alloys. In addition, the glass formed within a solid state may represent a distinct state of matter that may have other distinct properties yet to be explored, which may open new opportunities for research and development in the area of metallic glasses. □

Methods

Synchrotron X-ray experiments at high pressure and temperature were conducted on bulk, polycrystalline zirconium specimens of 1.0 mm diameter and 0.5 mm thickness, with the diffraction volume defined by 0.1×0.1 mm collimators. After observing the formation of amorphous zirconium, particular care was taken in our experiments, in that the formation of glassy zirconium was confirmed by collecting data at several different locations in the sample. The incident X-ray beam, however, is still too small compared to the bulk sample studied. In this regard, the GFA of bulk zirconium metal needs to be further studied by neutron diffraction.

In our high P – T neutron diffraction experiments, the cross-section of the incident neutron beam has a diameter of 5 mm, which is defined by cadmium and B_4C collimators. The time-of-flight neutron diffraction patterns were collected by eight detector banks that are available for TAP-98, at a fixed Bragg angle of $2\theta = 90^\circ$. The experiments were performed on the polycrystalline zirconium specimens of $\sim 100 \text{ mm}^3$ sample volume. This sample size, along with the current design of high-pressure cell, limits the experimental pressure to 5 GPa at high temperatures. So, unlike our X-ray diffraction experiments, our neutron diffraction experiments did not reach the P – T conditions needed for the glassy zirconium to form. These results are mainly used to constrain the α – β phase boundary, and will be published elsewhere.

Received 5 December 2003; accepted 4 June 2004; doi:10.1038/nature02715.

- Löffler, J. F. Bulk metallic glasses. *Intermetallics* **11**, 529–540 (2003).
- Johnson, W. L. Bulk glass-forming metallic alloys: science and technology. *MRS Bull.* **24**, 42–56 (1999).
- Inoue, A. Stabilization of metallic supercooled liquid and bulk amorphous alloys. *Acta Mater.* **48**, 279–306 (2000).
- Busch, R. The thermophysical properties of bulk metallic glass-forming liquids. *J. Miner. Metals Mater. Soc.* **52**, 39–42 (2000).
- Miller, M. K. Decomposition of bulk metallic glasses. *Mater. Sci. Eng. A* **250**, 133–140 (1998).
- Löffler, J. F. & Johnson, W. L. Model for decomposition and nano-crystallization of deeply undercooled $\text{Zr}_{41.2}\text{Ti}_{13.8}\text{Cu}_{12.5}\text{Ni}_{10}\text{Be}_{22.5}$. *Appl. Phys. Lett.* **76**, 3394–3396 (2000).
- Kundig, A. A., Löffler, J. F., Johnson, W. L., Uggowitzer, P. J. & Thiagarajan, P. Influence of decomposition on the thermal stability of undercooled Zr–Ti–Cu–Ni–Al alloys. *Scripta Mater.* **44**, 1269–1273 (2001).
- Busch, R., Schneider, S., Peker, A. & Johnson, W. L. Decomposition and primary crystallization in undercooled $\text{Zr}_{41.2}\text{Ti}_{13.8}\text{Cu}_{12.5}\text{Ni}_{10}\text{Be}_{22.5}$ melts. *Appl. Phys. Lett.* **67**, 1544–1546 (1995).
- Wang, W. H. *et al.* Role of addition in formation and properties of Zr-based bulk metallic glasses. *Intermetallics* **10**, 1249–1257 (2002).
- Eckert, J., Mattern, N., Zinkevitch, M. & Seidel, M. Crystallization behavior and phase formation in Zr–Al–Cu–Ni metallic glass containing oxygen. *Mater. Trans.* **39**, 623–632 (1998).
- Busch, R., Masuhr, A., Bakke, E. & Johnson, W. L. Bulk metallic glass formation from strong liquids. *Mater. Sci. Forum* **269**, 547–552 (1998).
- Ashby, M. F. & Jones, D. R. *Engineering Materials 2: An Introduction to Microstructures, Processing and Design* (Pergamon, Oxford, 1986).
- Weidner, D. J., *et al.* in *High-Pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) 13–17 (Geophysics Monograph Series, Vol. 67, American Geophysical Union, Washington DC, 1992).

- Vaughan, M. T. *et al.* T-cup: A new high-pressure apparatus for X-ray studies. *Rev. High Press. Sci. Technol.* **7**, 1520–1522 (1998).
- Zhao, Y., Von Dreele, R. B. & Morgan, J. G. A high P – T cell assembly for neutron diffraction up to 10 GPa and 1500 K. *High Press. Res.* **16**, 161–177 (1999).
- Decker, D. L. High-pressure equation of state for NaCl, KCl and CsCl. *J. Appl. Phys.* **42**, 3239–3244 (1971).
- Jayaraman, A., Klement, W. & Kennedy, G. C. Solid-solid transition in titanium and zirconium at high pressures. *Phys. Rev.* **131**, 644–649 (1963).
- Xia, H., Duclos, S. J., Ruoff, A. L. & Vohra, Y. K. New high-pressure phase transition in zirconium metal. *Phys. Rev. Lett.* **64**, 204–207 (1990).
- Xia, H., Ruoff, A. L. & Vohra, Y. K. Temperature-dependence of the ω –bcc phase transition in zirconium metal. *Phys. Rev. B* **44**, 10374–10376 (1991).
- Jona, F. & Marcus, P. M. Zirconium under pressure: Structure anomalies and phase transitions. *J. Phys. Condens. Matter* **15**, 5009–5019 (2003).
- Ostani, S. A. & Trubitsin, V. Y. Calculation of the P – T phase diagram of Zr in different approximations for the exchange-correlation energy. *Phys. Rev. B* **57**, 13485–13490 (1998).
- Hemley, R. J., Jephcoat, A. P., Mao, H. K., Ming, L. C. & Manghnani, M. H. Pressure-induced amorphization of crystalline silica. *Nature* **334**, 52–54 (1988).
- Mishima, O., Calvert, L. D. & Whalley, E. Melting of ice-I at 77 K and 10 Kbar: A new method of making amorphous solids. *Nature* **310**, 393–395 (1984).
- Highmore, R. J. & Greer, A. L. Eutectics and the formation of amorphous alloys. *Nature* **339**, 363–365 (1988).
- Richet, P. & Gillet, P. Pressure-induced amorphization of minerals: a review. *Eur. J. Miner.* **9**, 907–933 (1997).
- Heiming, A. *et al.* Phonon-dispersion of the bcc phase of group-IV metals 2: Bcc zirconium, a model case of dynamic precursors of martensitic transitions. *Phys. Rev. B* **43**, 10948–10962 (1991).
- Olijnyk, H. & Jephcoat, A. P. Effect of pressure on Raman phonons in zirconium metal. *Phys. Rev. B* **56**, 10751–10753 (1997).
- Pinsook, U. & Ackland, G. J. Calculation of anomalous phonons and the hcp–bcc phase transition in zirconium. *Phys. Rev. B* **59**, 13642–13649 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was performed under the auspices of the US Department of Energy with the University of California. The synchrotron X-ray diffraction experiments were conducted at X17B2 beamline of NSLS and BM-13 beamline of APS, operated by COMPRES and GESCARS, respectively.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.Z. (jzhang@lanl.gov).

Hydrological response to a seafloor spreading episode on the Juan de Fuca ridge

Earl Davis¹, Keir Becker², Robert Dziak³, John Cassidy¹, Kelin Wang¹ & Marvin Lilley⁴

¹Pacific Geoscience Centre, Geological Survey of Canada, Sidney, British Columbia V8L 4B2, Canada

²Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

³Oregon State University/NOAA, Hatfield Marine Science Centre, Newport, Oregon 97365, USA

⁴School of Oceanography, University of Washington, Seattle, Washington, 98195, USA

Seafloor hydrothermal systems are known to respond to seismic and magmatic activity along mid-ocean ridges, often resulting in locally positive changes in hydrothermal discharge rate, temperature and microbial activity, and shifts in composition occurring at the time of earthquake swarms and axial crustal dike injections^{1–10}. Corresponding regional effects have also been observed¹¹. Here we present observations of a hydrological response to seafloor spreading activity, which resulted in a negative formation-fluid pressure transient during and after an earthquake swarm in the sediment-sealed igneous crust of the

Middle Valley rift of the northernmost Juan de Fuca ridge. The observations were made with a borehole seal and hydrologic observatory originally established in 1991 to study the steady-state pressure and temperature conditions in this hydrothermally active area^{12,13}. The magnitude of the co-seismic response is consistent with the elastic strain that would be expected from the associated earthquakes, but the prolonged negative pressure transient after the swarm is surprising and suggests net co-seismic dilatation of the upper, permeable igneous crust. The rift valley was visited four weeks after the onset of the seismic activity, but no signature of increased hydrothermal activity was detected in the water column. It appears that water, not magma, filled the void left by this spreading episode.

A large seismic swarm began on 7 September 2001 in the northern part of the Middle Valley rift (Fig. 1). Initial activity was concentrated roughly 30 km north of the instrumented Ocean Drilling Program (ODP) Hole 857D, where the pressure observations presented here were made (Fig. 2). The area of activity

expanded to the south along the eastern side of the valley towards the borehole site over a period of roughly ten days (Fig. 3a), then continued for a period totalling about 20 days. The total slip energy, expressed in Fig. 3b as the cumulative seismic moment, is dominated by the largest events of the swarm and is concentrated in a four-day period of time beginning five days after the swarm onset. The mechanisms for the larger events indicate a range of motion from normal to strike-slip (Fig. 1), but all are consistent with the least-compressive stress in the valley being aligned with the direction of spreading.

Changes in formation pressure observed in the borehole display a clear correlation with the seismic activity (Fig. 3b). Pressure begins to decline gradually by a few tenths of a kilopascal over the initial five days of the swarm, then drops in three discrete steps at the times of the largest earthquakes. These stepwise decreases in pressure probably reflect the elastic response of the formation in the immediate vicinity of the borehole to distant fault slip, with the size of each step (1.0, 0.8 and 2.7 kPa) being controlled by a combination of the magnitude ($M_w = 6.0, 5.5$ and 5.7 , respectively) and proximity to the borehole (28, 18 and 14 km, respectively) of the corresponding earthquake. The magnitudes of local elastic strain estimated from the pressure steps are 0.3×10^{-6} , 0.2×10^{-6} and 0.8×10^{-6} , respectively (see Methods section). Given the uncertainties in the earthquake locations and the position of the borehole relative to the nodal planes (where strain gradients are large), it is difficult to put these observations properly into the

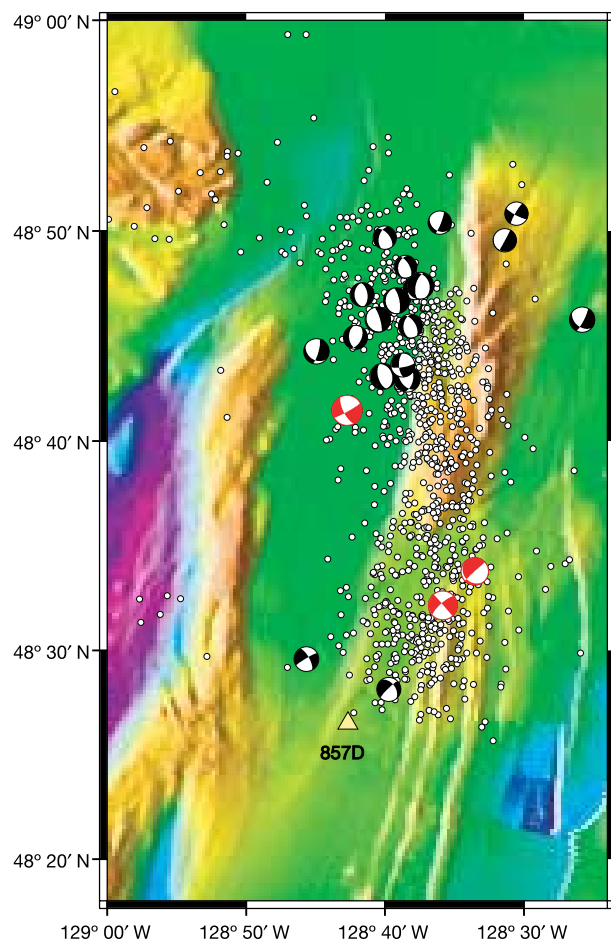


Figure 1 Locations of events of the September 2001 seismic swarm in Middle Valley, northern Juan de Fuca Ridge, determined from offshore SoSUS hydroacoustic array tertiary-phase data (error less than 4 km)²², along with earthquake locations determined using Western Canadian Seismic Network data (<http://www.pgc.nrcan.gc.ca/seismo>) and mechanisms (compressional quadrants shown solid) constrained by moment tensor solutions (J. Risteanu, personal communication). Earthquakes that produced the discrete pressure steps seen in Fig. 2 are shown in red. ODP Hole 857D, located at the southern limit of the seismicity, was drilled and instrumented with a CORK borehole observatory in September 1991 to allow natural formation temperatures and pressures to be determined after drilling perturbations had dissipated^{12,13}. The hole is lined with solid steel casing through unconsolidated sediments of the rift to a depth of 574 m below the seafloor (m.b.s.f.), and extends as an open hole in permeable basement to a total depth of 936 m.b.s.f.

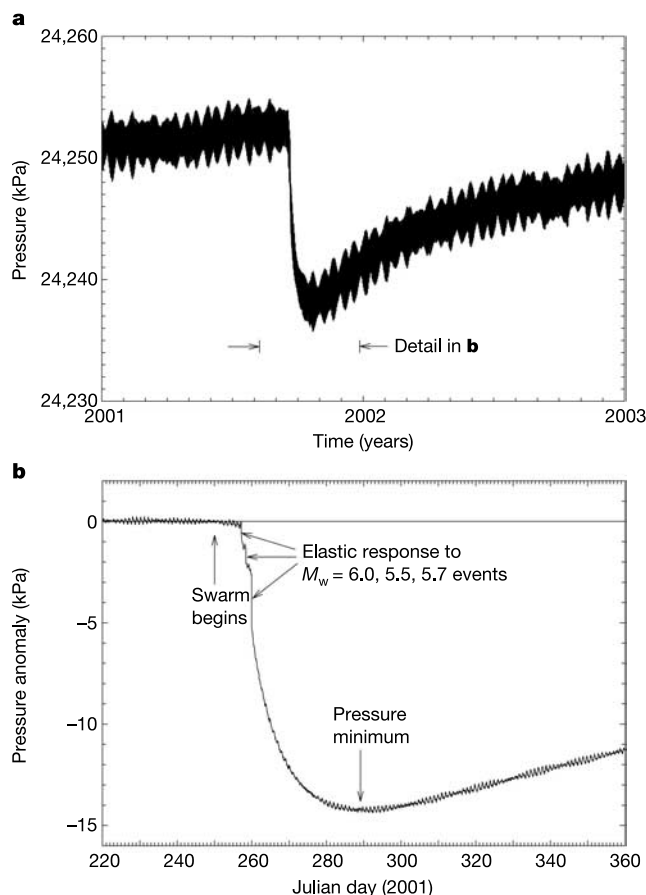


Figure 2 Formation pressure measured hourly in ODP Hole 857D before and after the September 2001 seismic swarm. Pressures measured at the sea floor (not shown) reflect the variable load applied to the formation by the water column and atmosphere; those measured below the CORK seal at the top of the casing string reflect the average formation pressure over the open-hole interval. The total formation pressure is shown in **a**, and the anomalous pressure transient with the effects of seafloor loading removed is shown in **b**.

context of a predicted co-seismic elastic strain field, as has been done in other instances of co-seismic observations^{11,14}, but they do fall well within the range predicted at the site for the corresponding earthquakes using simple elastic dislocation theory¹⁵.

Following the co-seismic elastic response, pressure continued to decrease for 30 days, reaching a minimum 14.3 kPa below the pre-swarm level (9.7 kPa below the level observed immediately after the greatest moment release; Fig. 2b), before beginning a year-long recovery towards an unperturbed state (Fig. 2a). The decrease following the swarm may reflect continuing dilatational elastic strain at the borehole site, although given the relatively large size of that portion of the anomaly and the very small seismic moment accumulated after the largest events (Fig. 3b), this would require virtually all of any later strain to have occurred without earthquakes. A more likely explanation is that the continued decline reflects lateral hydraulic diffusion through the sediment-sealed igneous crust (basement) from the earthquake source area to the borehole.

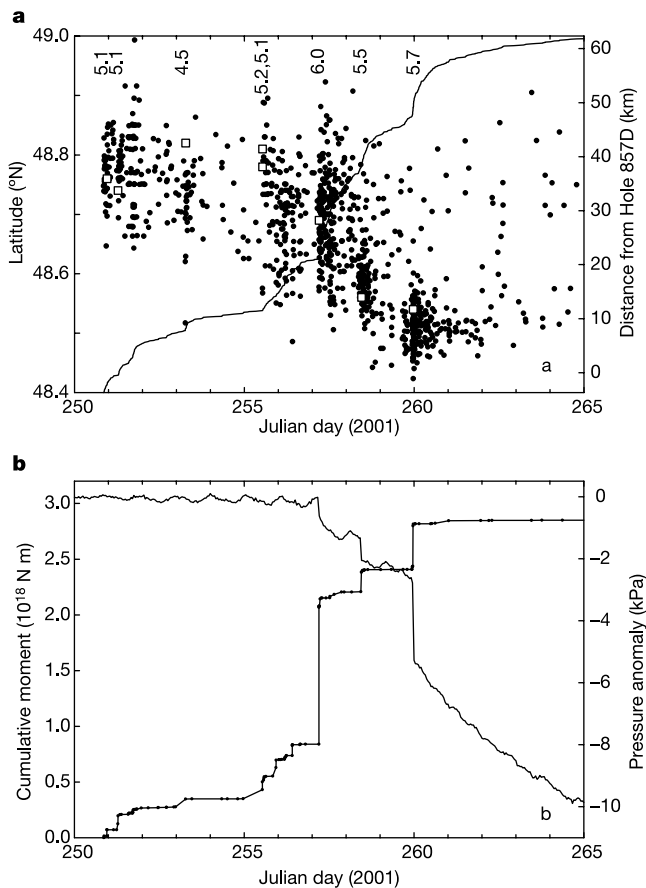


Figure 3 Simultaneous seismic and hydrologic records of the 2001 Middle Valley earthquake swarm. Hydroacoustic (solid points) and seismic (open squares) event locations (given as distance north of ODP Hole 857D) are plotted as a function of time in **a**. The line showing the cumulative number of hydroacoustic events (totalling over 1,000 over the course of the swarm) reveals an increase and subsequent decay of the frequency of events after each large earthquake. The cumulative seismic moment (**b**, line with points) is calculated from moment magnitudes for the largest events and from body-wave magnitudes converted to moment magnitudes for intermediate events. An approximate relationship between acoustic amplitude and seismic magnitude suggests that inclusion of hydroacoustic events might add 10–20% to the total cumulative moment. Formation pressure (descending line in **b**) is shown as an anomaly relative to the average background pressure before the seismic swarm. Effects of oceanographic and atmospheric loading (see Fig. 2a) have been removed; the residual diurnal/semi-diurnal signal seen in **b** and in Fig. 2b reflects the contribution of earth tides or a diffusive (non-zero phase) ocean tidal loading component.

A simple model is considered here to explore the cause of the post-seismic pressure transient and to place a crude constraint on the hydraulic properties of the formation. The earthquake swarm is treated as a concentrated source of dilatancy located 30 km from the observation hole, with diffusion occurring axisymmetrically about a vertical line source in a fully confined permeable layer. This is directly analogous to a hydrologic ‘slug test’¹⁶ in which a sudden fluid injection or withdrawal in one borehole is monitored in an observation hole some distance away. Characteristic curves for the propagation of a pressure pulse past an observation point are shown in Fig. 4. Predicted pressures at a given range reach minima at times that depend on the hydraulic diffusivity η ; given the geometry assumed, the observed 30-day transit time (Fig. 2b) suggests $\eta \approx 90 \text{ m}^2 \text{ s}^{-1}$. Given reasonable formation and fluid physical properties (see Methods section), this lateral formation-scale diffusivity corresponds to a permeability of $3 \times 10^{-12} \text{ m}^2$. This value is generally consistent with borehole pumping test results completed at the time of drilling in the valley¹⁷, and with results of numerical convection simulations¹⁸ which require high basement permeability to yield the degree of uniformity of upper basement temperatures inferred from seafloor heat flux and seismic reflection data¹⁹. A crude estimate for the ‘effective dilatation’ associated with the source of the transient can be estimated with this same model by considering the volume of the slug of water that would need to be withdrawn at the source location to produce the magnitude of the observed pressure transient (that is, by comparing the predicted transient in Fig. 4 with the observed transient in Fig. 2). With the hydraulic diffusivity—and thus permeability and transmissivity—constrained by the diffusion time, and the maximum pressure anomaly of roughly 10 kPa (relative to the co-seismic elastic pressure anomaly) observed at a range of 30 km, a value for the ratio of the withdrawal volume (or in this application, the new pore volume created) to the layer thickness is estimated to be $V/b = 2 \times 10^4 \text{ m}^3 \text{ m}^{-1}$. If this withdrawal volume is generated by dilatation along the 30-km-long portion of the axis where seismic activity is concentrated (Fig. 3), an effective ‘opening’ of 0.6 m is suggested. Although the numbers are not directly related, this amount of opening is comparable to the magnitude of fault slip implied by the swarm’s cumulative moment, $\sim 3 \times 10^{18} \text{ N m}$. A

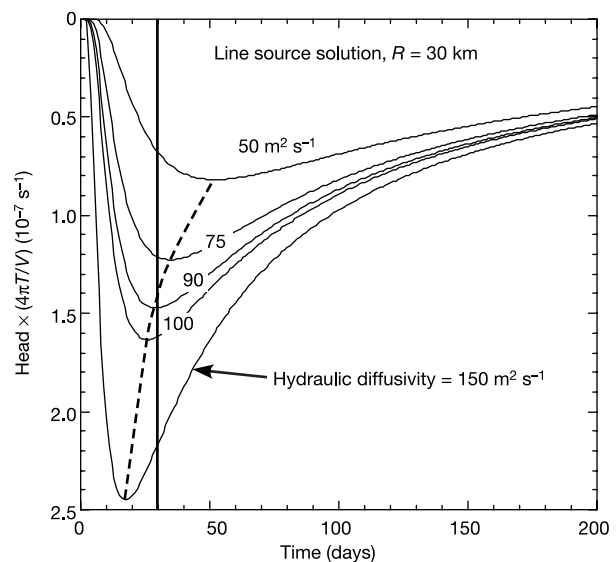


Figure 4 Head resulting from an impulse propagating axisymmetrically in a uniform porous medium, calculated at a range of 30 km using a line-source approximation for a hydrologic slug test¹⁶, which is valid at large distances from the source well. In the plot, head, $H = P/\rho g$, is scaled by the hydraulic transmissivity, $T = bk\rho g/\mu$ (where b is layer thickness, k is permeability, ρ is fluid density, and μ is viscosity) and by the slug volume, V (see Methods for a summary of properties).

value of roughly 1 m is suggested if the area of rupture extended over the 3-km thickness of the crust that has been observed to host microseismicity²⁰, and along the 30-km length of the axis defined by the concentration of seismicity seen in Figs 1 and 3. A value of 0.5 m is suggested if rupture penetrated the full thickness of the crust.

Although the general character of the predicted 'slug-test' transient (Fig. 4) is remarkably similar to that of the observed transient (Fig. 2), the results must be considered with appropriate caution, because there are large uncertainties in applying such a simple model. For example, (1) the model considers only lateral diffusion. Ventilation through the sediment section or through faults and permeable basement outcrops (as reflected by the year-long recovery to the normal formation state, established by the thermal and hydrologic structure of the valley¹³) will 'dilute' the magnitude of the initial transient, and vertical diffusion from any extension of the dilatant volume below the transmissive uppermost crust will add to and stretch the source signal. (2) Considering the dilatation to be concentrated as a vertical line-source 'implosion' is unquestionably to oversimplify. Distributed dilatancy along the axis of the valley will create diffusional directivity (enhancing propagation across strike), and reduce the effective range between the source and the observation site along axis. (3) Perhaps the greatest shortcoming of the slug-test analogue is that it does not consider the initial distribution of elastic strain associated with the spreading event, in which quadrants of contraction (and hence volumes of positive anomalous pressure) lie next to quadrants of dilatation. Diffusion between these quadrants will produce an 'annihilation' effect near nodal planes and complicate any pressure transient generated by net dilatation. Despite these uncertainties, however, the simple concentrated-source model provides a useful first-order tool for estimating net dilatation and the hydraulic transmission properties of the formation, and for guiding a more complete analysis that will take distributed co-seismic deformation and three-dimensional diffusion into account.

We note that the co- and post-seismic pressure transients observed in ODP Hole 857D represent a significant fraction (totalling nearly 20%) of the deep-seated buoyancy that drives hydrothermal discharge at a nearby vent field. A pressure differential of +80 kPa was measured across the sediment section in ODP Hole 858G, which was drilled 1.6 km to the north into a sediment-buried permeable basement edifice beneath the vent field, sealed, and instrumented at the same time as ODP Hole 857D^{12,13}. Unfortunately, the seals at this high-temperature site failed long before the earthquake swarm, but it is reasonable to conclude that basement there experienced a similar co- and post-seismic reduction in pressure. The lack of any signs of augmented hydrothermal discharge searched for during a water-column investigation of the valley carried out shortly after the swarm ended is consistent with this conclusion, as well as with our primary conclusion that the permeable crust at this site suffered net dilatancy at the time of the seismic activity. □

Methods

An outline of the relationships among strain, total stress, and pore fluid pressure in a poroelastic medium is provided in refs 11 and 21. The elastic properties involved are either established by laboratory data (at conditions corresponding to the formation temperature and pressure of 280 °C and 30 MPa, respectively), by observations of cored material, or by observed formation-fluid pressure response to tidal loading. These include: fluid compressibility, $\beta_f = 1.3 \times 10^{-9} \text{ Pa}^{-1}$; solid constituent compressibility, $\beta_s = 2.0 \times 10^{-11} \text{ Pa}^{-1}$; one-dimensional tidal loading efficiency, $\gamma = 0.14$; matrix frame compressibility, $\beta_m = 6.8 \times 10^{-11} \text{ Pa}^{-1}$; porosity, $n = 0.15$; and Poisson's ratio, $\nu = 0.25$. The resulting linear coefficient relating strain and fluid pressure is $0.29 \times 10^{-9} \text{ Pa}^{-1}$. The relationship between permeability k and hydraulic diffusivity η involves a subset of these properties plus viscosity $\mu = 10^{-4} \text{ Pa s}$ (appropriate for the formation temperature): $k = \mu \xi \eta$, where ξ , the storage compressibility, is a function of n, β_f, β_s and β_m as outlined in ref. 21.

Received 15 January; accepted 14 June 2004; doi:10.1038/nature02755.

1. Baker, E. T. *et al.* Hydrothermal event plumes from the CoAxial seafloor eruption site, Juan de Fuca Ridge. *Geophys. Res. Lett.* **22**, 151–154 (1995).

2. Embley, R. W., Chadwick, W. W., Jonasson, I. R., Butterfield, D. A. & Baker, E. T. Initial results of the rapid response to the 1993 CoAxial event: Relationships between hydrothermal and volcanic processes. *Geophys. Res. Lett.* **22**, 143–146 (1995).
3. Von Damm, K. L. *et al.* Evolution of East Pacific Rise hydrothermal vent fluids following a volcanic eruption. *Nature* **395**, 47–50 (1995).
4. Dziak, R. P., Fox, C. G. & Schreiner, A. E. The June–July 1993 seismoacoustic event at CoAxial Segment, Juan de Fuca Ridge: Evidence for a lateral dike injection. *Geophys. Res. Lett.* **22**, 135–138 (1995).
5. Sohn, R. A., Fornari, D. J., Von Damm, K. L., Hildebrand, J. A. & Webb, S. C. Seismic and hydrothermal evidence of a cracking event on the East Pacific Rise crest near 9° 50' N. *Nature* **396**, 159–161 (1998).
6. Delaney, J. R. *et al.* The quantum event of ocean crustal accretion: Impacts of diking at mid-ocean ridges. *Science* **281**, 222–230 (1998).
7. Dziak, R. P. & Fox, C. G. The January 1998 earthquake swarm at Axial Volcano, Juan de Fuca Ridge: Hydroacoustic evidence of seafloor volcanic activity. *Geophys. Res. Lett.* **26**, 3429–3432 (1999).
8. Lupton, J. E., Baker, E. T. & Massoth, G. J. Helium, heat, and the generation of hydrothermal event plumes at mid-ocean ridges. *Earth Planet. Sci. Lett.* **171**, 343–350 (1999).
9. Johnson, H. P. *et al.* Earthquake-induced changes in a hydrothermal system at the Endeavour Segment, Juan de Fuca Ridge. *Nature* **407**, 174–177 (2000).
10. Lilley, M. D., Butterfield, D. A., Lupton, J. E. & Olson, E. J. Magmatic events can produce rapid changes in hydrothermal vent chemistry. *Nature* **422**, 878–881 (2003).
11. Davis, E. E., Wang, K., Thomson, R. E., Becker, K. & Cassidy, J. F. An episode of seafloor spreading and associated plate deformation inferred from crustal fluid pressure transients. *J. Geophys. Res.* **106**, 21953–21963 (2001).
12. Davis, E. E., Becker, K., Pettigrew, T., Carson, B. & Macdonald, R. CORK: A hydrologic seal and downhole observatory for deep-ocean boreholes. *Proc. ODP Init. Rep.* **139**, 43–53 (1992).
13. Davis, E. E. & Becker, K. Formation temperatures and pressures in a sedimented rift hydrothermal system: Ten months of CORK observations, Holes 857D and 858G. *Proc. ODP Sci. Res.* **139**, 649–666 (1994).
14. Jonsson, S., Segall, P., Pedersen, R. & Björnsson, G. Post-earthquake ground movements correlated to pore-pressure transients. *Nature* **424**, 179–183 (2003).
15. Okada, Y. Internal deformation due to shear and tensile faults in a half space. *Bull. Seismol. Soc. Am.* **82**, 1018–1040 (1992).
16. Cooper, H. H. Jr, Bredehoeft, J. D. & Papadopoulos, I. S. Response of a finite diameter well to an instantaneous charge of water. *Wat. Resour. Res.* **3**, 263–269 (1967).
17. Becker, K., Morin, R. H. & Davis, E. E. Permeabilities in the Middle Valley hydrothermal system measured with packer and flowmeter experiments. *Proc. ODP Sci. Res.* **139**, 613–626 (1994).
18. Bessler, J. U., Smith, L. & Davis, E. E. Hydrologic and thermal conditions at a sediment/basalt interface: Implications for interpretation of field measurements at Middle Valley. *Proc. ODP Sci. Res.* **139**, 667–678 (1994).
19. Davis, E. E. & Villinger, H. Tectonic and thermal structure of the Middle Valley sedimented rift, northern Juan de Fuca Ridge. *Proc. ODP Init. Rep.* **139**, 9–41 (1992).
20. Golden, C. E., Webb, S. C. & Sohn, R. A. Hydrothermal microearthquake swarms beneath active vents at Middle Valley, northern Juan de Fuca Ridge. *J. Geophys. Res.* **108**, doi:10.1029/2001JB000226 (2003).
21. Wang, K. Applying fundamental principles and mathematical models to understand processes and estimate parameters. In *Hydrogeology of the Oceanic Lithosphere* (eds Davis, E. E. & Elderfield, H.) (Cambridge Univ. Press, in the press).
22. Fox, C. G., Dziak, R. P., Marumoto, H. & Schreiner, A. E. Potential for monitoring low-level seismicity on the Juan de Fuca Ridge using military hydrophone arrays. *Mar. Technol. Soc. J.* **27**, 22–30 (1994).

Acknowledgements We acknowledge the Ocean Drilling Program for ongoing support for long-term borehole monitoring experiments, T. Pettigrew, R. Meldrum, and R. Macdonald for engineering assistance, R. Thomson and S. Mihaley for assistance with the post-swarm water-column survey on board CHS Tully, M. Fowler for assistance with SoSUS data processing, and J. Ristau for earthquake moment tensor calculations. Financial support was provided by the US National Science Foundation and by the Geological Survey of Canada.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.E.D. (edavis@nrcan.gc.ca).

Cladogenesis and morphological diversification in passerine birds

Robert E. Ricklefs

Department of Biology, University of Missouri-St Louis, 8001 Natural Bridge Road, St Louis, Missouri 63121-4499, USA

Morphological diversity tends to increase within evolving lineages over time^{1,2}, but the relative roles of gradual evolutionary change (anagenesis)³ and abrupt shifts associated with speciation events (cladogenesis, or 'punctuated equilibrium')⁴ have not been resolved for most groups of organisms⁵. However,

these two modes of evolution can be distinguished by the fact that morphological variance increases in proportion to time under anagenesis⁶, and in proportion to the logarithm of the number of species under cladogenesis⁷. Although species and time are themselves correlated, multiple regression analysis provides a statistical framework for partitioning their relative contributions. In this study, I use multiple regressions to evaluate the effects of time and species number on morphological diversity within clades of passerine birds. The results show clearly that number of species exerts a strong influence on morphological variance independent of time, but that time has no unique effect. Thus, morphological evolution in birds seems to be associated with cladogenesis. How lineage splitting promotes morphological diversification poses an important challenge to ecologists and evolutionary biologists.

I quantified morphological variation by the variance of principal component (PC) scores within tribe-to-family-level clades of passerine birds (Supplementary Table 1) based on eight log-transformed external measurements (for example, wing, leg and beak; see Methods). These variables characterize structure related to diet, substrate use and habitat^{8,9}. The variances provide convenient indices to diversity because the PC axes have dimensionless scales, distances between species in morphological space are conserved and the estimate of the variance is independent of both sample size and the average value of the measurement variables^{1,10}.

I evaluated the contributions of taxon age (see Methods) and species richness to morphological variation by multiple regression. Using this approach, the unique statistical effects of age and species richness can be estimated, independently of any correlation between the two. Morphological variance within the entire sample of clades ($n = 95$) was unrelated to variation in the relative age of clades ($P > 0.18$ for all PC axes), but increased significantly ($P < 0.05$) with respect to the logarithm of species on all principal components, except PC4, PC5 and PC8 (Fig. 1a). When age was dropped from the analysis and continental versus non-continental distribution was included as an effect, morphological variances on PC4 ($P = 0.0003$) and PC8 ($P = 0.006$) were also significantly related to species number (Fig. 1b).

In many clades of birds distributed in southeast Asia, including Indonesia, the Philippines and Australasia, numerous differentiated island populations have been given species names. To avoid potentially uninformative variation introduced by these clades, I performed a second analysis restricted to continental clades, with undersampled taxa also removed (see Methods). Among these 50 clades, relative age remained insignificant in both simple and

multiple regressions. When age was removed as a variable, the relationship of morphological variance to logarithm of the number of species was significant for all principal components ($P < 0.0001$; $P = 0.0049$ (PC5) and $P = 0.0008$ (PC8)), except PC7 ($P = 0.19$). Number of species alone explained 15.3% to 43.6% of the variation among clades on individual PC axes, and 33.3% of the total variation. These results were independent of the proportion of species sampled per clade (all $P > 0.05$).

Because the relative ages of clades varied by only a factor of four (see Methods), these analyses might not have included a sufficient range to detect an effect of time on morphological variance. Indeed, age and the log-transformed number of species per clade were unrelated in the entire sample of clades ($r = -0.135$, $P = 0.19$, $n = 95$), and morphological variance was unrelated to age alone in simple correlations (all $P > 0.12$). To provide a broader range of clade age, I calculated morphological variances within 108 genera (having five or more species sampled) and included these in a multiple regression with tribe-to-family-level clades. In the analysis, morphological variance was significantly related to age alone on all PC axes ($P < 0.01$); this was as expected because the logarithm of species number was also significantly correlated with relative age ($r = 0.55$, $P < 0.0001$). However, the effect of age in multiple regressions was insignificant for all axes ($P > 0.05$) except PC5 ($P = 0.016$). As found before, the effect of the logarithm of species number was significant ($P < 0.006$ for all PC axes, and < 0.0001 for PC1, 2, 3 and 5), accounting for 9.1% to 24.1% of the morphological variance. Thus, morphological variance does not increase over time independently of species production, and I conclude that the proliferation of species is the primary driver of morphological diversification within clades of passerine birds.

In spite of the significant relationship between morphological diversity and species richness, considerable variation remains. Factors besides species formation *per se* that could influence morphological diversity include: (1) the complexity of the environment as a template for morphological adaptation^{11–13}; (2) the constraining influence of other species, particularly potential competitors, within the distribution of a diversifying clade; and (3) diversifying selection from competition by secondarily sympatric sister taxa^{11,14–17}.

The environmental template undoubtedly is clade-specific and difficult to evaluate. Although large and small clades coexist regionally, whether the availability of ecological space limits diversification cannot be easily addressed. One might expect islands to present more limited ecological opportunities than continental

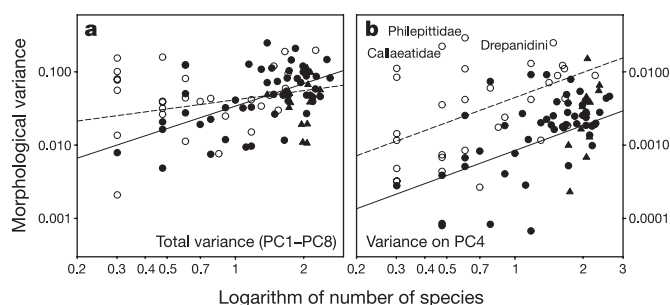


Figure 1 The relationship between morphological variance and number of species for 95 passerine clades. Clades distributed in non-continental (open symbols) and continental (solid symbols) regions, and undersampled clades (triangles). **a**, Total variance (PC1–PC8). Dashed regression line includes all 95 clades (slope = 0.415 ± 0.141 s.e.m.); solid regression line includes only well-sampled continental clades ($n = 50$, slope = 1.016 ± 0.208 s.e.m.). **b**, PC4 (relative beak length). Dashed regression line for non-continental clades (elevation = 0.723 ± 0.180 s.e.m.); solid regression line for all continental clades (common slope = 1.134 ± 0.306 s.e.m.).

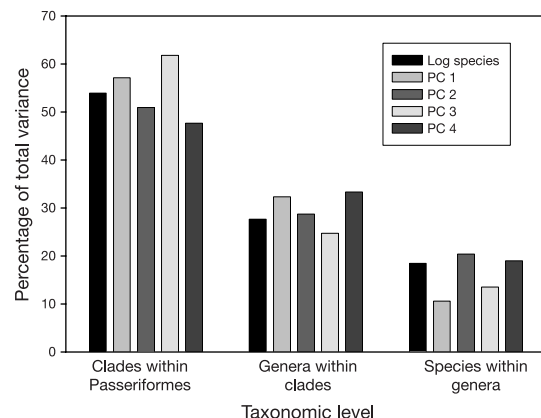


Figure 2 Distribution of morphological variance on PC axes 1–4. Values estimated by hierarchical nested analysis of variance compared to the value expected if morphology diversified in direct relation to the logarithm of the average number of species within evolutionary lineages (black bars).

landmasses; however, even most non-continental clades have diversified within large, ecologically complex regions. As explained below, non-continental clades of comparable size have no less morphological variance than those that have diversified within continental landmasses.

To evaluate the constraining influence of regional diversity, I compared morphological variance between clades in continental and non-continental areas. Many non-continental tribe-to-family-level lineages of passerine birds occur on islands or island-continents (including New Guinea, New Zealand, Madagascar and Australia) where morphological space presumably has been less densely occupied than in the major continental landmasses of Africa, Eurasia, and North and South America. Variance within continental clades was less than that in island clades on all morphological axes (binomial $P = 0.004$). However, the only axis showing a significant continent/non-continent effect in the overall analysis ($n = 95$) was PC4, which is correlated with the length of the beak relative to the size of the foot. Variance on this axis in non-continental clades exceeded that in continental clades ($F_{1,92} = 16.1$, $P < 0.0001$) by 0.72 ± 0.18 (mean $\pm$ s.e.m.) $\log_{10}$ units, or a factor of 5.2 (Fig. 1b). This axis is associated with variation in diet, feeding method and substrate use in birds, and it probably represents an axis along which birds can easily diverge ecologically.

A range of beak morphology, from greatly elongated to massively stout, characterizes many island passerine clades. Three of these with extreme variation on PC4 are indicated in Fig. 1b: the New Zealand wattlebirds (Callaeidae), the Malagasy asities (Philepittidae) and the Hawaiian honeycreepers (Drepanidini). These unusual radiations of non-continental forms might reflect open ecological space on islands allowing novel avenues of morphological change. Indeed, the beak morphology of the Hawaiian honeycreepers fills most of the morphological space occupied by passerine birds¹⁸.

To test the effect of diversifying selection, I further related morphological variation to the extent of sympatry (see Methods) among species within small clades. Diversifying selection should be most intense when sister lineages gain secondary sympatry and compete strongly for shared resources^{15,17,19}. The potential effect of both factors was explored for clades with fewer than ten species ($n = 32$), that exhibited the full range of morphological variance (see Fig. 1) and could be scored easily for degree of sympatry (see Supplementary Table 2), and also could be designated as either continental or non-continental.

Morphological variances were higher among non-continental clades on all axes except PC7 (binomial $P = 0.035$), but the effects were not significant by themselves on any of the PC axes. Variance increased with respect to the sympatry index on all eight PC axes ($P = 0.004$), but significantly only on PC4 ($P = 0.014$). Thus, among small clades, morphological diversification tends to be accelerated on islands and by selection following secondary sympatry, but neither effect is strong.

Whatever the cause of diversification, if morphological change associated with speciation events were random and homogeneous, the relationship between morphological variance and the logarithm of species number would be linear. Accordingly, the slope of the regression between the logarithm of the variance and the double logarithm of species number would be 1.0. Higher slopes would indicate acceleration of morphological diversification with species number; lower slopes would indicate constrained diversification with increasing species richness. Among the well-sampled continental clades ($n = 50$), the slope (b) of this relationship exceeded one for PC1 ($b = 1.74 \pm 0.34$ s.e.m., $t = 2.18$, $P = 0.034$) and PC3 ($b = 1.54 \pm 0.26$ s.e.m., $t = 2.22$, $P = 0.034$). The total variance in the smaller PC axes increased at a lower rate than the logarithm of species number (PC5–8 combined, $b = 0.49 \pm 0.15$ s.e.m., $t = -3.40$, $P = 0.0029$). Thus, some components of morphological variance increase disproportionately more rapidly than the number of species per clade, but these are balanced by some that increase more

slowly. The resulting total morphological variance (PC1–8) increased as the 1.02 ± 0.21 s.e.m. power of the logarithm of species number.

This direct proportionality of morphological diversification with the logarithm of the number of species extends to the entire sample of passerine birds. Total morphological variance among all passerines (clade size = 5,712 species) was 111% of the extrapolated values for the sum of PC1–8 (based on the regression of variance on species number among well-sampled continental clades). Thus, diversification does not seem to be constrained by existing diversity.

The distribution of morphological variance among taxonomic levels provides an additional indication that diversification is unconstrained by increasing species number. If variance increased uniformly with the logarithm of species number, it would be distributed as 53.9% among 106 tribe-to-family-level clades within Passeriformes, 27.7% among an average of 10.95 genera within clades and 18.4% among an average of 4.9 species within genera. Morphological variance corresponds closely to this expectation, as shown for the first four PC axes in Fig. 2. The hierarchical distribution of variance argues that recent lineage splitting, represented by the species-within-genus component of variance, has resulted in continued morphological diversification along all axes of variation.

The filling of morphological space probably has not constrained speciation either. Tribe-to-family-level clades show no evidence of regulation of species richness in terms of the expected variation in number of species among clades²⁰. That is, large tribe-to-family-level clades are not fewer than predicted by the geometric distribution produced by a random speciation-extinction model. Diversification of passerine birds with respect to both species richness and morphology seems to be unrestrained globally. Whether or not increasing species number results in finer local partitioning of geography or habitat that is unrelated to morphology, larger clades also generate greater morphological diversity.

Morphological diversification in passerine birds is strongly associated with species rather than time *per se*. However, the strong species effect on morphological diversity identified in this analysis does not necessarily imply punctuated equilibrium—change associated with the speciation process itself²¹. Speciation can also promote evolutionary diversification indirectly by establishing selection for divergent evolution in evolutionarily independent lineages. Thus, divergent evolution might occur gradually over long periods following lineage splitting rather than as an abrupt consequence of the formation of new species in small, peripherally isolated populations forced to undergo rapid reorganization of their gene pools²². Although secondary sympatry of sister species is likely to create strong selection for divergence^{15,16,19}, newly formed lineages in species-rich clades might also face sufficient selective pressure from more distantly related species, to diversify morphologically in allopatry. Mechanisms linking morphological diversity to species richness can be investigated more readily by relating divergence to age, distribution and ecology in studies focused on individual clades. □

Methods

Taxonomy

This analysis is limited to passerine birds (the monophyletic order Passeriformes, 5,712 species), which include more than half of all bird species. I use the 106 smallest non-nested suprageneric taxa designated as monophyletic groups (tribes, subfamilies or families) in the treatment of Sibley and Monroe²³ (see Supplementary Table 1). Age was estimated as the depth of the node uniting each group with its sister taxon, obtained by DNA hybridization and expressed as the difference in temperature at 50% dissociation (the melting point temperature, ΔT_{50H}) between heteroduplexed and homoduplexed DNA ($^{\circ}\text{C}$)²⁴. Relative ages of the 106 taxa ranged between 4.3 and 17.9 $^{\circ}\text{C}$ and number of species varied between 1 and 413. Number of species was independent of deeper phylogenetic relationships among clades²⁴ according to Abouheif's test²⁵ of serial phylogenetic independence ($n = 106$, $P = 0.155$).

Measurements and samples

The following measurements were obtained with calipers and rulers from museum skins: total length, wing and tail (in mm); tarsus, middle toe, and the length, width and breadth

of the culmen (0.1 mm)¹⁰. Measurements were averaged within species. Altogether, the data set comprised 1,018 (17.8% of 5,712) species in 103 (97.2% of 106) suprageneric taxa of passerine birds.

Two samples of tribe-to-family-level clades were analysed. The most inclusive contained the 95 taxa for which I measured more than one species and could therefore calculate morphological variance. The second sample ($n = 50$) excluded clades restricted to non-continental landmasses (primarily Australasia, New Zealand, Madagascar and the Greater Antilles), as well as tribe-to-family-level clades for which I measured fewer species than one-third the number of genera or one-tenth the number of species.

In a further analysis, I extended the range of clade age by including estimates of morphological variance within 108 genera having five or more species and for which I had measured more than 20% of all species. Genera were assigned a relative age of $\Delta T_{50H} = 3.2^\circ\text{C}$ (the median relative age of nodes within subtribes)²⁴. These were combined with well-sampled tribe-to-family-level clades.

Sympatry within small clades

I used distribution maps in field guides and handbooks to estimate by eye the approximate degree of sympatry among the species in clades having fewer than ten species (see Supplementary Table 2). For each clade, the sympatry index is the proportion of species (0–1) having greater than $\approx 50\%$ overlap in their geographic range with one or more other species in the clade, hence having the potential for interacting locally.

Analyses

The eight morphological values were $\log_{10}$ -transformed to make the distribution of variation in each of the measurements dimensionless, approximately normal within the sample of species and unrelated to the mean value of the measurement. Over all species, standard deviations of the log-transformed measurements varied between 0.144 and 0.195 (equivalent to factors of 1.39 and 1.57) and thus large and small measurements contributed comparably to distance between species in morphological space. Log-transformed measurements were subjected to a principal components (PC) analysis based on the covariance matrix, to reduce the dimensionality of the data and obtain uncorrelated axes of morphological variation. Using the covariance matrix preserves the original euclidean distances between species²⁶.

The first PC axis incorporated 75% of the variance in morphology and the first three axes incorporated 91%. The resulting eight PC scores for each species were then used to estimate the variance on each of the PC axes for each tribe-to-family-level clade. The estimate of the variance is unbiased with respect to sample size¹. Among 95 clades, the morphological standard deviation was phylogenetically independent according to Abouheif's test²⁵ for PC1 ($P = 0.55$) and PC4 ($P = 0.17$).

To test the hypothesis that morphological diversification is a function of number of species, time or both, I used multiple regression to determine the relationship between morphological variance on each of the PC axes and the logarithm of species number and relative age of each clade. Each of the variables was log-transformed to make the deviations of morphological variance about the regression uniform and to test the linearity of the relationship (slope = 1).

I conducted nested analyses of variance to examine the distribution of variation among species within genera, genera within tribe-to-family-level clades, and clades within passerines. The expected morphological variation depending on a direct relationship to the logarithm of species was calculated as follows: the logarithm of the number of smaller taxa within each larger taxon at each level in the hierarchy of analysis (that is, 106 clades within passerines, 10.95 genera per clade and 4.92 species per genus).

All analyses were performed using the Statistical Analysis System (SAS Institute, Cary, North Carolina).

Received 3 November 2003; accepted 24 May 2004; doi:10.1038/nature02700.

1. Foote, M. The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**, 129–152 (1997).
2. Harmon, L. J., Schulte, J. A. II, Larson, A. & Losos, J. B. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* **301**, 961–964 (2003).
3. Simpson, G. G. *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
4. Gould, S. J. & Eldredge, N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* **3**, 115–151 (1977).
5. Jackson, J. B. C. & Cheetham, A. H. Tempo and mode of speciation in the sea. *Trends Ecol. Evol.* **14**, 72–77 (1999).
6. Raup, D. M. & Gould, S. J. Stochastic simulation and evolution of morphology—towards a nomothetic paleontology. *Syst. Zool.* **23**, 305–322 (1974).
7. Foote, M. in *Evolutionary Paleobiology* (eds Jablonski, D., Erwin, D. H. & Lipps, J. H.) 62–86 (Univ. Chicago Press, Chicago, 1996).
8. Miles, D. B. & Ricklefs, R. E. The correlation between ecology and morphology in deciduous forest passerine birds. *Ecology* **65**, 1629–1640 (1984).
9. Ricklefs, R. E. & Miles, D. B. in *Ecological Morphology. Integrative Organismal Biology* (eds Wainwright, P. C. & Reilly, S. M.) 13–41 (Univ. Chicago Press, Chicago, 1994).
10. Ricklefs, R. E. & Travis, J. A morphological approach to the study of avian community organization. *Auk* **97**, 321–338 (1980).
11. Benkman, C. W. Divergent selection drives the adaptive radiation of crossbills. *Evolution* **57**, 1176–1181 (2003).
12. Moritz, C., Patton, J. L., Schneider, C. J. & Smith, T. B. Diversification of rainforest faunas: an integrated molecular approach. *Annu. Rev. Ecol. Syst.* **31**, 533–563 (2000).
13. Schluter, D. Morphological and phylogenetic relations among the Darwin's finches. *Evolution* **38**, 921–930 (1984).
14. Lack, D. *Darwin's Finches* (Cambridge Univ. Press, Cambridge, 1947).
15. Grant, P. R. Convergent and divergent character displacement. *Biol. J. Linn. Soc.* **4**, 39–68 (1972).
16. Schluter, D. Character displacement and the adaptive divergence of finches on islands and continents. *Am. Nat.* **131**, 799–824 (1988).

17. Schluter, D. Ecological causes of adaptive radiation. *Am. Nat.* **148**, S40–S64 (1996).
18. Lovette, I. J., Bermingham, E. & Ricklefs, R. E. Clade-specific morphological diversification and adaptive radiation in Hawaiian songbirds. *Proc. R. Soc. Lond. B* **269**, 37–42 (2002).
19. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
20. Ricklefs, R. E. Global diversification rates of passerine birds. *Proc. R. Soc. Lond. B* **270**, 2285–2291 (2003).
21. Eldredge, N. & Gould, S. J. in *Models in Paleobiology* (ed. Schopf, T. J. M.) 82–115 (Freeman, San Francisco, 1972).
22. Carson, H. L. & Templeton, A. R. Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annu. Rev. Ecol. Syst.* **15**, 97–131 (1984).
23. Sibley, C. G. & Monroe, B. L. Jr. *Distribution and Taxonomy of Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
24. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of the Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
25. Abouheif, E. A method to test the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* **1**, 895–909 (1999).
26. Legendre, P. & Legendre, L. *Numerical Ecology* (Elsevier, Amsterdam, 1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements I thank M. Foote, T. Givnish and S. S. Renner for constructive comments. For access to specimens, I am grateful to curators and collection managers at the Academy of Natural Sciences in Philadelphia, the US National Museum, the American Museum of Natural History, the Field Museum in Chicago, the Bavarian State Collections in Munich, the Senckenberg Museum in Frankfurt and the Museum des Sciences Naturelles in Brussels. This work was supported in part by research funds available to Curators' Professors at the University of Missouri.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.R. (ricklefs@umsl.edu).

The combined effects of pathogens and predators on insect outbreaks

Greg Dwyer¹, Jonathan Dushoff² & Susan Harrell Yee¹

¹Department of Ecology and Evolution, 1101 E 57th Street, University of Chicago, Chicago, Illinois 60637-1573, USA

²Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA

The economic damage caused by episodic outbreaks of forest-defoliating insects has spurred much research¹, yet why such outbreaks occur remains unclear². Theoretical biologists argue that outbreaks are driven by specialist pathogens or parasitoids, because host–pathogen and host–parasitoid models show large-amplitude, long-period cycles resembling time series of outbreaks^{3,4}. Field biologists counter that outbreaks occur when generalist predators fail, because predation in low-density defoliator populations is usually high enough to prevent outbreaks^{5–8}. Neither explanation is sufficient, however, because the time between outbreaks in the data is far more variable than in host–pathogen and host–parasitoid models^{1,2}, and far shorter than in generalist-predator models^{9–11}. Here we show that insect outbreaks can be explained by a model that includes both a generalist predator and a specialist pathogen. In this host–pathogen–predator model, stochasticity causes defoliator densities to fluctuate erratically between an equilibrium maintained by the predator, and cycles driven by the pathogen^{12,13}. Outbreaks in this model occur at long but irregular intervals, matching the data. Our results suggest that explanations of insect outbreaks must go beyond classical models to consider interactions among multiple species.

The host–pathogen model that we begin with describes the effects of a specialist pathogen on the population dynamics of a forest defoliator. Host-specific pathogens of defoliators are often baculoviruses, which cause fatal diseases transmitted when larvae

consume foliage contaminated with infectious cadavers¹⁴. Although baculoviruses infect only larvae, multiple rounds of transmission within a generation can lead to high infection rates, and so epidemics are often severe in high-density defoliator populations^{1,15,16}. Infected larvae are converted into infectious cadavers, and pathogens are reintroduced into defoliator populations the following year when hatchlings encounter infectious particles from the previous year¹⁷. Previous work has shown that these dynamics can be accurately predicted by a standard epidemic model, modified to allow for variability in susceptibility among hosts¹⁸. The fraction of infected hosts, I , in such a model can be closely approximated by

the following implicit expression, derived by assuming that the epidemic proceeds until no further infections occur¹⁹:

$$1 - I(N_t, Z_t) = \left(1 + \frac{\bar{v}}{\mu k} (N_t I(N_t, Z_t) + \rho Z_t) \right)^{-k} \quad (1)$$

Here N_t and Z_t are host and pathogen densities in generation t , μ is the rate at which cadavers lose infectiousness, and ρ is the susceptibility of hatchlings relative to later-stage larvae. To allow for variability in susceptibility among larvae, transmission rates in the model follow a gamma distribution with average transmission $\bar{v}$ and inverse squared coefficient of variation k . Because most outbreaking defoliators have only one generation per year²⁰, generations are discrete. Between model generations, surviving defoliators lay eggs, and virus particles must survive in the environment until new host larvae hatch. The host–pathogen model is therefore

$$N_{t+1} = \lambda N_t (1 - I(N_t, Z_t)) \quad (2)$$

$$Z_{t+1} = f N_t I(N_t, Z_t) \quad (3)$$

Here λ is net defoliator fecundity, f is pathogen over-winter survival, and I is calculated using equation (1). For realistic parameter values, this model shows long-period, large-amplitude cycles in defoliator densities¹⁹. The period and amplitude of these cycles are similar to those seen in time series for many outbreaking insects¹⁹, except that the cycle period is much more regular in the model than in the data. To see whether this discrepancy could be explained through the addition of stochasticity, we multiplied the right-hand side of equation (2) by a log-normally distributed random variable ε_t , with median 1. The model cycles, however, are so robust that stochasticity has only a modest effect (Fig. 1a).

We therefore modified the model to allow for a generalist predator, because low-density defoliator populations often experience severe mortality from generalists such as birds⁷, spiders⁵, small mammals⁶ or generalist parasitoids^{8,21}. Generalists by definition rely on multiple resources, and so, in contrast to specialists, their densities respond weakly or not at all to changes in prey density. A generalist's attack rate can nevertheless respond to density, if, for example, the generalist switches among prey items according to the relative availability of each item²². Allowing for such behaviour

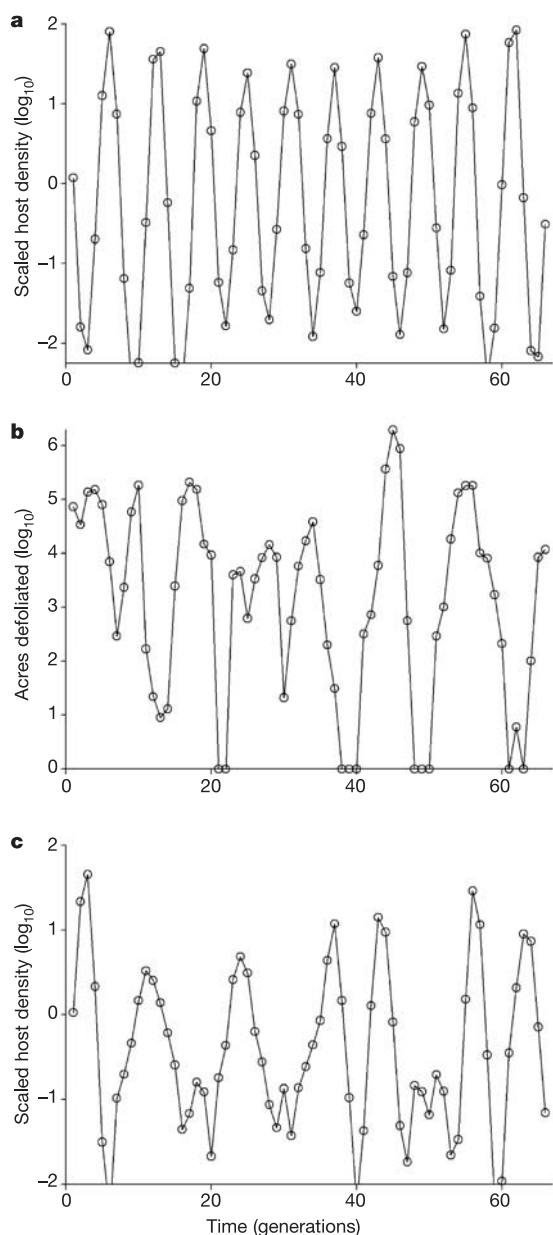


Figure 1 Comparison of model output with data for an outbreaking insect. **a**, Host–pathogen model, equations (1)–(3). **b**, Gypsy moth defoliation in New Hampshire, USA. **c**, Combined model, equations (1), (4) and (5). Parameter values are in the Methods, with the addition that the standard deviation of $\log_{10}$ of the forcing term ε_t is 0.5. Model output is scaled as described in the Methods. Note that the data give areas defoliated, whereas the models give densities, so in comparing models with data we focus on the time between outbreaks rather than on outbreak amplitudes.

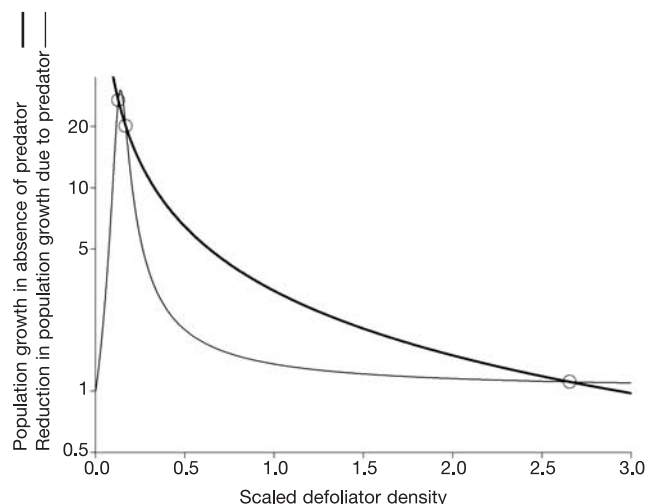


Figure 2 Graphical representation of the combined model's equilibria. The dark line is the defoliator's quasi-equilibrium population growth in the absence of the predator, and the light line is the relative reduction in the defoliator growth rate due to the predator. Circled intersections are equilibria.

produces the model:

$$N_{t+1} = \lambda N_t (1 - I(N_t, Z_t)) \left(1 - \frac{2abN_t}{b^2 + N_t^2} \right) \quad (4)$$

$$Z_{t+1} = fN_t I(N_t, Z_t) \quad (5)$$

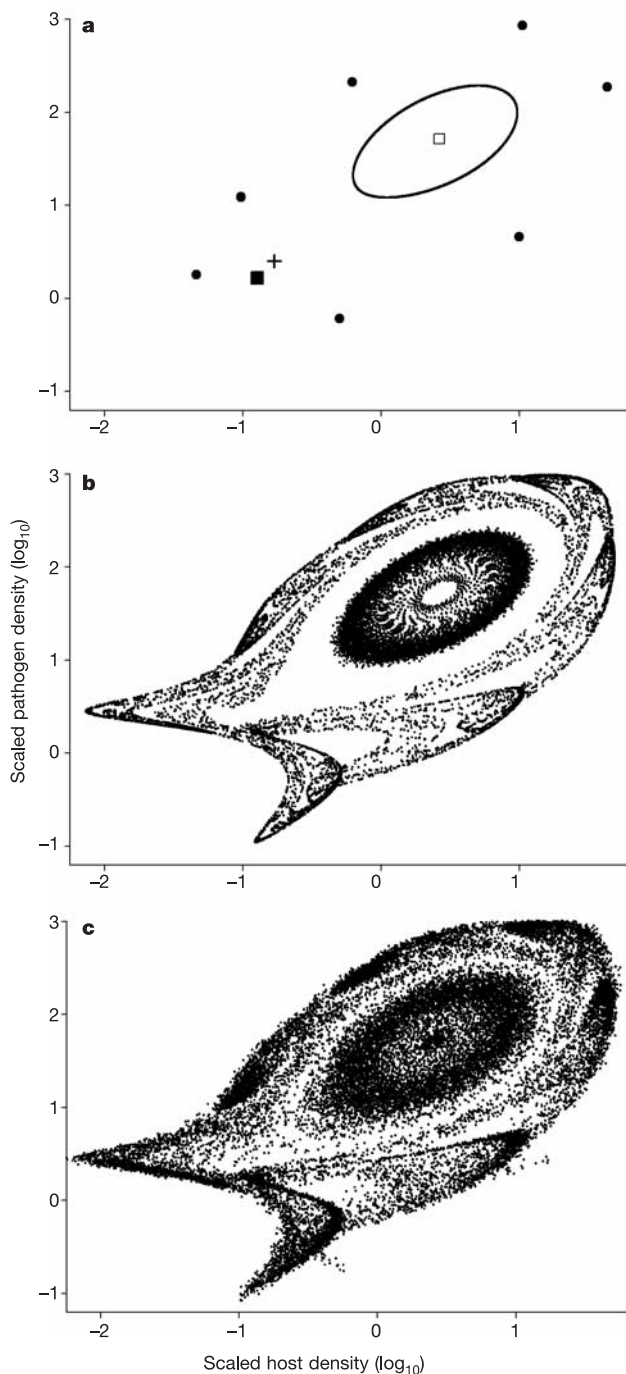


Figure 3 Phase portraits of the combined model, with time proceeding anticlockwise. **a**, Deterministic attractors. The ellipse is a quasi-periodic attractor and the small circles are a phase-locked limit cycle. Source, sink and saddle-point equilibria are depicted by an open square, a closed square and a cross, respectively. **b**, Short-term dynamics for many initial conditions. We iterated the model for 250 generations, discarded the first 150 generations, plotted values for the remaining generations¹², and repeated for initial N_t from 0.001 to 500 and initial Z_t from 0.01 to 10,000, multiplying by 1.4 at each step. **c**, Long-term dynamics with stochasticity. The standard deviation of the $\log_e$ of the forcing term is 0.05. Note that, unlike in **b**, densities never settle on a single attractor.

Here the fraction of defoliators killed by the predator is $2abN_t/(b^2 + N_t^2)$, where a is the maximum fraction killed, and b is the defoliator density at which the fraction killed is maximized. Note that the fraction killed by the predator rises rapidly with increasing prey density as the predator specializes on the newly abundant prey. At high densities, however, the predator is overwhelmed, and the attack rate levels off. The reduction in the defoliator's population growth rate due to the predator is therefore maximized at an intermediate density (Fig. 2). Generalist predators usually have little impact at high density, but often impose density-dependent regulation on low-density defoliator populations^{5–8,21}, consistent with this model.

The addition of the generalist predator leads to dynamics that are much more complex than the dynamics of the host–pathogen model. This dynamical complexity arises in part because the combined model can have multiple equilibria. To find these equilibria, we consider a quasi-equilibrium¹¹ at which pathogen density instantaneously adjusts to host density. At this quasi-equilibrium, we can calculate the defoliator's population growth rate as a function of defoliator density alone (Fig. 2). Model equilibria then occur at densities for which the quasi-equilibrium population growth rate in the absence of the predator is equal to the reduction due to the predator (Fig. 2). The model can have multiple equilibria because the reduction due to the predator reaches a maximum at an intermediate density.

Our results hold for a wide range of parameter values (see Supplementary Information), but to describe the model's dynamics in more detail, we use values calculated for the gypsy moth (*Lymantria dispar*) in North America (for data sources see the Methods). For these values, the combined model has a high-density equilibrium at which the defoliator is controlled by the pathogen and the predator is relatively unimportant, much as in the host–pathogen model (Fig. 2). In contrast to the host–pathogen model, however, there is also a low-density equilibrium at which these roles are reversed, so that the predator controls the defoliator and the pathogen is relatively unimportant. This predator-maintained equilibrium is stable, but only locally, and the equilibrium coexists with two attractors associated with the high-density equilibrium (Fig. 3a). The long-term model dynamics therefore depend on the initial densities of defoliators and pathogens; moreover, many initial conditions lead to complex transient dynamics that can last many generations¹² (Fig. 3b). The occurrence of multiple attractors and long transients in turn means that adding a small amount of stochasticity—by multiplying the right-hand side of equation (4) by the log-normal random variable ε_t —causes trajectories to move unpredictably among attractors¹³ (Fig. 3c). For slightly different parameter values the low-density equilibrium becomes unstable and the system exhibits deterministic chaos. These chaotic dynamics are qualitatively similar to the stochastically induced complex dynamics shown in Fig. 3c.

This dynamical complexity is important because it can help to explain the high levels of variability in outbreak interval seen in

Table 1 Analysis of time between outbreaks for some forest-defoliating insects

Species	Location	Average period (yr)	Coefficient of variation of period
<i>Bupalus piniarius</i> ²⁸	Germany	9.8	0.44
<i>Epirrita autumnata</i> ²⁹	Finland	9.5	0.26
<i>Lymantria dispar</i> ²⁷	Maine, USA	9.3	0.19
	Massachusetts, USA	10.5	0.59
	New Hampshire, USA	8.1	0.37
	Vermont, USA	10.0	0.67
	British Columbia, Canada	9.3	0.22
<i>Orygia pseudotsugata</i> ¹⁶	Washington, USA	13.0	0.62
	N Idaho and NE Oregon, USA	9.0	0.16
	SW Idaho, USA	10.8	0.30
	Japan	14.6	0.43
<i>Quadralcarifera punctatella</i> ³⁰			

most (but not all²) long-term defoliator time series (Fig. 1, Table 1). Direct comparison of model output with data is problematic, however, because the model is highly sensitive to initial conditions, and we do not have even crude estimates of the starting densities for any defoliator. One solution is to compare statistical moments of the model output with statistical moments of the data²³, and so we compare the average and the coefficient of variation of the time between outbreaks in the model with those in the data. Figure 4 shows that the combined model yields long average times between outbreaks, and high levels of variability in the time between outbreaks. Although confidence intervals are wide, the mean and the coefficient of variation of the time between outbreaks in the combined model are at least consistent with most of the data. In contrast, although the host–pathogen model—equations (1)–(3)—also shows a long average time between outbreaks, its cycles are very robust, and so achieving similar levels of variability in that model requires very high levels of stochasticity.

Because of the importance of weather in defoliator population dynamics², we have assumed that the major source of stochasticity is random fluctuations in parameter values caused by extrinsic events, so-called ‘environmental stochasticity’. Although ‘demographic

stochasticity’, the chance events that befall small populations, could conceivably be important in low-density defoliator populations, most such populations are of large spatial extent, and so we suspect that absolute numbers of most defoliators generally do not fall low enough for demographic stochasticity to be of great importance. A related issue is that forest defoliator populations tend to be synchronized over large areas², but for some chaotic ecological models, the addition of spatial structure can lead to asynchrony among subpopulations²⁴. When we add spatial structure to our model, however, we still see high levels of synchrony among subpopulations (see Supplementary Information). Understanding why our model gives spatial synchrony when other models do not requires additional research, but we suspect that an important part of the explanation is that the long time between outbreaks in our model makes it more difficult for subpopulations to get out of phase with one another. Our results are similarly robust to changes in the specialist natural enemy. If we instead begin with a host–parasitoid model, adding a generalist predator again gives multiple equilibria, and complex dynamics that arise in response to stochasticity (see Supplementary Information).

Our results show that the addition of a generalist predator to a classical host–pathogen model can create a stable, low-density equilibrium, and that interactions between this equilibrium and limit cycles induced by the pathogen lead to stochastically induced complex dynamics, and thus high variability in the time between insect outbreaks. These dynamics differ not only from those of classical host–pathogen and host–parasitoid models, but also from those of classical generalist–predator models. Although classical generalist–predator models can also have multiple equilibria^{9–11}, they assume that high-density defoliator populations are kept in check by competition for resources. Outbreaks in these models are separated by decades of low, stable defoliator densities, and so biologists have assumed that outbreaks will only occur when generalist predators fail because of weather or other stochastic factors^{1,5–8}. In contrast, because we realistically assume that high-density populations crash owing to a specialist pathogen, the upper equilibrium in our model shows high-amplitude cycles. The stabilizing effect of the generalist predator in our model is therefore much smaller than in classical generalist–predator models. Our work suggests that two-species models are insufficient for understanding outbreaks, whether in insects or in other outbreaking animal taxa⁸, and that classical theories of outbreaks must be extended to consider interactions among multiple species. Moreover, models with multiple equilibria are generally used only to describe catastrophic shifts in ecosystems²⁵. Our work suggests that an additional important effect of multiple equilibria is the creation of complex dynamics. □

Methods

To reduce the number of parameters, we rewrite the model equations (1), (4) and (5) using the non-dimensionalized host and pathogen densities $\tilde{N}_t \equiv \mu N_t / \bar{\nu}$ and $\tilde{Z}_t \equiv \mu Z_t / \bar{\nu}$ (ref. 19). This gives the rescaled equations

$$1 - I(N_t, Z_t) = \left(1 + \frac{1}{k}(N_t I(N_t, Z_t) + Z_t)\right)^{-k} \quad (6)$$

$$N_{t+1} = \lambda N_t (1 - I(N_t, Z_t)) \left(1 - \frac{2a\hat{b}N_t}{\hat{b}^2 + N_t^2}\right) \quad (7)$$

$$Z_{t+1} = \phi N_t I(N_t, Z_t) \quad (8)$$

where $\hat{b}$ is the ratio of the density at maximum predation b to the epidemic threshold $N_e \equiv \mu / \bar{\nu}$, while ϕ is the between-season impact of the pathogen. At the quasi-equilibrium, pathogen density adjusts to its equilibrium value instantaneously: that is, we set pathogen density Z_t to its equilibrium value $\tilde{Z} = \phi N_t I(N_t, \tilde{Z})$. This allows us to eliminate Z_t from equation (6), so that equilibria occur at the intersections of the functions $f(N_t) \equiv \lambda(1 - I(N_t))$ and $g(N_t) \equiv 1/[1 - 2a\hat{b}N_t/(\hat{b}^2 + N_t^2)]$.

Our estimates of the transmission parameters $\bar{\nu}$, k and μ are averages from field experiments with the gypsy moth baculovirus¹⁹. Studies of low-density gypsy moth populations permit estimation of the reproductive rate, λ , independently of density-dependent sources of mortality⁶, and provide an estimate of the equilibrium density during non-outbreak years. Although this is not quite the same as b , the density at

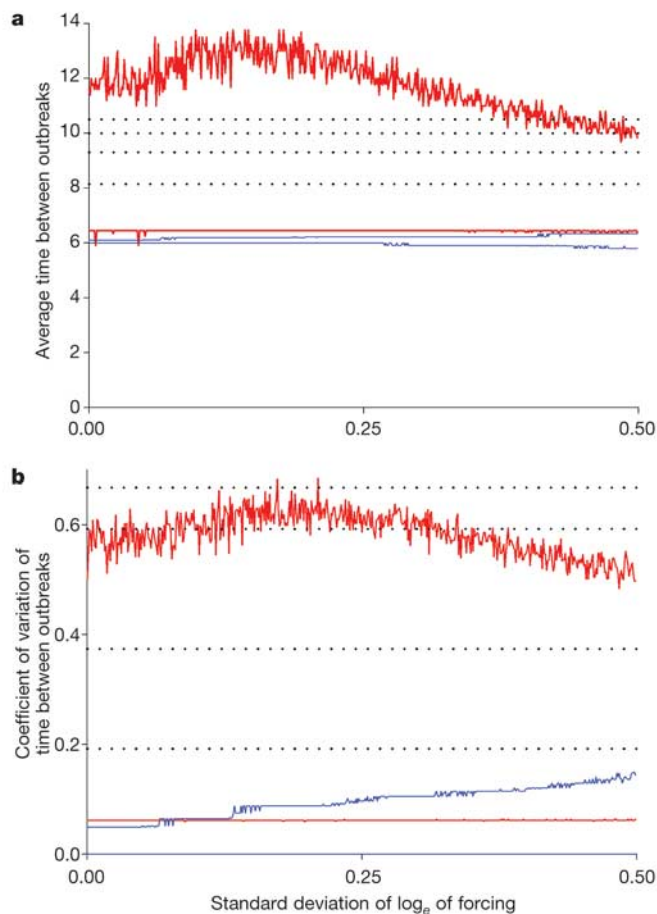


Figure 4 Effects of stochasticity on time between outbreaks. **a**, Average time between outbreaks. **b**, Coefficient of variation of time between outbreaks. We used the models to generate 2,000 realizations of 134 generations each, the approximate number of generations for which the gypsy moth has been resident in North America. We then calculated the average and the coefficient of variation of the time between outbreaks for the final 65 generations, the length of the gypsy moth time series used in Table 1. Initial densities of hosts and pathogens were drawn randomly from a uniform distribution between 0.01 and 100. Lines depict 95th centiles of each statistic. Red lines are for the combined model equations (1), (4) and (5). Blue lines are for the host–pathogen model equations (1)–(3). Dotted lines indicate data for the gypsy moth (Table 1).

maximum predation, the two densities are close relative to the uncertainty in our estimates. These calculations give the following values: $\lambda = 74.6$; $k = 1.06$; $b = 0.14$.

Estimates of predation rates in gypsy moth populations often exceed 99% (ref. 21), which for our parameter values gives deterministic chaos in the combined model in the absence of stochasticity. Because these experimental estimates may be slightly inflated relative to predation rates in natural populations⁶, in the figures we use the slightly lower value of $a = 0.967$. For $a = 0.967$ and $\phi = 20$, the model has three equilibria, with the lowest one stable (the intermediate-density equilibrium is always unstable), usefully illustrating the origins of complex dynamics in our model. In fact, as we demonstrate in the Supplementary Information, our results hold for a wide range of parameter values. We estimated the pathogen over-winter survival parameter, ϕ , by first estimating the other parameters, and then adjusting ϕ until the amplitude of density fluctuations in each model matched estimates of the amplitude of density fluctuations derived from the literature²⁶ (note that the data used to estimate ϕ are densities, and are thus unrelated to the areas defoliated in Figs 1 and 4). Deriving an estimate of the density amplitude from the literature requires that we make some assumption about the detection threshold, the lowest density that can be detected. For detection thresholds of 1–2 egg masses per hectare, our amplitude estimates range from 3.38 to 3.68 orders of magnitude. In this range of amplitude estimates, the inherent stochasticity of the combined model makes it difficult to estimate ϕ with more than one significant digit. Given these uncertainties, we take 3.5 orders of magnitude as our estimate of the observed amplitude, which gives best-fitting values $\phi = 60$ for the combined model and $\phi = 100$ for the host–pathogen model. In Fig. 3, however, we use $\phi = 20$ to illustrate the origins of complex dynamics.

In calculating the statistics in Table 1, we restricted ourselves to time series with at least five outbreaks, to reduce the uncertainty in our estimates of the coefficient of variation of the time between outbreaks. Also, to illustrate irregularity in the inter-outbreak period, we used only species for which the coefficient of variation of the time between outbreaks was at least 0.15. In the case of the gypsy moth (*Lymantria dispar*) and the Douglas-fir tussock moth (*Orgyia pseudotsugata*) the data are spatially referenced. Because the model assumes that populations are well-mixed—so that dispersal does not limit species interactions—for the gypsy moth we used data for particular states, while for the tussock moth we used outbreak regions as in the original source¹⁶. At these scales, populations are nearly synchronous²⁷. Data were periods of outbreaks, except for the following. *Lymantria dispar* data were acres defoliated, and outbreaks were defined as periods in which the area defoliated was more than 1,000 acres, a common definition of an outbreak for defoliation data. *Bupalus piniarius* data were densities, and outbreaks were defined as periods during which the insect's density was greater than its mean density.

Received 10 December 2003; accepted 13 April 2004; doi:10.1038/nature02569.

- Myers, J. H. Can a general hypothesis explain population cycles of forest Lepidoptera? *Adv. Ecol. Res.* **18**, 179–242 (1988).
- Liebold, A. & Kamata, N. Are population cycles and spatial synchrony a universal characteristic of forest insect populations? *Popul. Ecol.* **42**, 205–209 (2000).
- Varley, G. C., Gradwell, G. R. & Hassell, M. P. *Insect Population Ecology: An Analytical Approach* 135–153 (Blackwell Scientific, Oxford, 1973).
- Anderson, R. M. & May, R. M. The population-dynamics of micro-parasites and their invertebrate hosts. *Phil. Trans. R. Soc. Lond. B* **291**, 451–524 (1981).
- Mason, R. R., Torgerson, T. R., Wickman, B. E. & Paul, H. G. Natural regulation of a Douglas-fir tussock moth (Lepidoptera: Lymantriidae) population in the Sierra Nevada. *Environ. Entomol.* **12**, 587–594 (1983).
- Elkinton, J. S. *et al.* Interactions among gypsy moths, white-footed mice, and acorns. *Ecology* **77**, 2332–2342 (1996).
- Parry, D., Spence, J. R. & Volney, W. J. A. Response of natural enemies to experimentally increased populations of the forest tent caterpillar, *Malacosoma disstria*. *Ecol. Entomol.* **22**, 97–108 (1997).
- Klemola, T., Tanhuanpää, M., Korpimäki, E. & Ruohomäki, K. Specialist and generalist natural enemies as an explanation for geographical gradients in population cycles of northern herbivores. *Oikos* **99**, 83–94 (2002).
- Southwood, T. R. E. & Comins, H. N. A synoptic population model. *J. Anim. Ecol.* **45**, 949–965 (1976).
- May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
- Ludwig, D., Jones, D. D. & Holling, C. S. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. *J. Anim. Ecol.* **47**, 315–332 (1978).
- Rand, D. & Wilson, H. B. Chaotic stochasticity: a ubiquitous source of unpredictability in epidemics. *Proc. R. Soc. Lond. B* **246**, 179–184 (1991).
- Dennis, B., Desharnais, R. A., Cushing, J. M., Henson, S. M. & Constantino, R. F. Estimating chaos and complex dynamics in an insect population. *Ecol. Monogr.* **71**, 277–303 (2001).
- Cory, J. S., Hails, R. S. & Sait, S. M. In *The Baculoviruses* (ed. Miller, L. K.) 301–339 (Plenum, New York, 1997).
- Woods, S. & Elkinton, J. S. Bimodal patterns of mortality from nuclear polyhedrosis virus in gypsy moth (*Lymantria dispar*) populations. *J. Invertebr. Pathol.* **50**, 151–157 (1987).
- Shepherd, R. F. Evidence of synchronized cycles in outbreak patterns of Douglas-fir tussock moth, *Orgyia pseudotsugata* (McDunnough) (Lepidoptera: Lymantriidae). *Mem. Entomol. Soc. Can.* **146**, 107–121 (1988).
- Murray, K. D. & Elkinton, J. S. Environmental contamination of egg masses as a major component of transgenerational transmission of gypsy-moth nuclear polyhedrosis virus (LdMNPV). *J. Invertebr. Pathol.* **53**, 324–334 (1989).
- Dwyer, G., Elkinton, J. S. & Buonaccorsi, J. P. Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. *Am. Nat.* **150**, 685–707 (1997).
- Dwyer, G., Dushoff, J., Elkinton, J. S. & Levin, S. A. Pathogen-driven outbreaks in forest defoliators revisited: building models from experimental data. *Am. Nat.* **156**, 105–120 (2000).
- Hunter, A. F. In *Population Dynamics: New Approaches and Synthesis* (eds Cappuccino, N. & Price, P. W.) 41–64 (Academic, New York, 1995).
- Gould, J. R., Elkinton, J. S. & Wallner, W. E. Density-dependent suppression of experimentally created

- gypsy moth *Lymantria dispar* (Lepidoptera: Lymantriidae) populations by natural enemies. *J. Anim. Ecol.* **59**, 213–233 (1990).
- Holling, C. S. Some characteristics of simple types of predation and parasitism. *Can. Entomol.* **91**, 293–320 (1959).
- Kendall, B. E. *et al.* Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. *Ecology* **80**, 1789–1805 (1999).
- Bjornstad, O. N. Cycles and synchrony: two historical experiments and one 'experience'. *J. Anim. Ecol.* **69**, 869–873 (2000).
- Scheffer, S., Carpenter, S., Foley, J. A., Folkes, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
- Williams, D. W. *et al.* Oak defoliation and population density relationships for the Gypsy Moth (Lepidoptera: Lymantriidae). *J. Econ. Entomol.* **84**, 1508–1514 (1991).
- Williams, D. W. & Liebhold, A. M. Influence of weather on the synchrony of gypsy moth (Lepidoptera: Lymantriidae) outbreaks in New England. *Environ. Entomol.* **24**, 987–995 (1995).
- Schwerdtfeger, R. Über die Ursachen des Massenwechsels der Insekten. *Z. Angew. Entomol.* **28**, 254–303 (1941).
- Ruohomäki, K. *et al.* Causes of cyclicity of *Epirrita autumnata* (Lepidoptera, Geometridae): grandiose theory and tedious practice. *Popul. Ecol.* **42**, 211–223 (2000).
- Liebold, A., Kamata, N. & Jacobs, T. Cyclicity and synchrony of historical outbreaks of the beech caterpillar, *Quadralcarifera punctatella* (Motschulsky) in Japan. *Res. Popul. Ecol.* **38**, 87–94 (1996).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank O. Bjornstad, P. Turchin and A. Hunter for comments. G.D. and J.D. were supported by grants from the US National Science Foundation. J.D. was also supported by the Andrew W. Mellon Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.D. (gdwyer@uchicago.edu).

An SCF-like ubiquitin ligase complex that controls presynaptic differentiation

Edward H. Liao¹, Wesley Hung¹, Benjamin Abrams² & Mei Zhen¹

¹Department of Medical Genetics and Microbiology, Samuel Lunenfeld Research Institute, University of Toronto, Ontario, Canada M5G 1X5

²Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California 95064, USA

During synapse formation, specialized subcellular structures develop at synaptic junctions in a tightly regulated fashion. Cross-signalling initiated by ephrins, Wnts and transforming growth factor- β family members between presynaptic and post-synaptic termini are proposed to govern synapse formation^{1–3}. It is not well understood how multiple signals are integrated and regulated by developing synaptic termini to control synaptic differentiation. Here we report the identification of FSN-1, a novel F-box protein that is required in presynaptic neurons for the restriction and/or maturation of synapses in *Caenorhabditis elegans*. Many F-box proteins are target recognition subunits of SCF (Skp, Cullin, F-box) ubiquitin-ligase complexes^{4–7}. *fsn-1* functions in the same pathway as *rpm-1*, a gene encoding a large protein with RING finger domains^{8,9}. FSN-1 physically associates with RPM-1 and the *C. elegans* homologues of SKP1 and Cullin to form a new type of SCF complex at presynaptic periaxonal zones. We provide evidence that T10H9.2, which encodes the *C. elegans* receptor tyrosine kinase ALK (anaplastic lymphoma kinase¹⁰), may be a target or a downstream effector through which FSN-1 stabilizes synapse formation. This neuron-specific, SCF-like complex therefore provides a localized signal to attenuate presynaptic differentiation.

Drosophila and *C. elegans* provide genetic models for uncovering conserved regulatory mechanisms for synapse differentiation^{11,12}.

Studies indicate that the periaxial zone, a presynaptic region that is excluded from active zones and associated vesicles, might regulate synapse growth¹³. Mutations in RPM-1, a *C. elegans* periaxial zone protein, cause enlargement in some GABAergic neuromuscular junctions (NMJs) with an excessive number of active zones, whereas other GABAergic NMJs disappear. RPM-1 might restrict the development or promote the maturation of different synapses^{8,9}. Its *Drosophila* homologue, HIW, restricts synaptic growth at glutamatergic NMJs^{14,15}. The human homologue PAM¹⁶ is expressed in the nervous system. RPM-1/HIW/PAM family proteins contain a Ran-type guanine nucleotide exchange factor domain, multiple RING finger motifs, and a region with unknown structures. RING finger domains are found in many E3 ubiquitin ligases as docking sites for ubiquitin-conjugating enzymes¹⁷, indicating that RPM-1/HIW/PAM might regulate synapse development through a ubiquitin-dependent process. This is supported by the genetic interaction between *hiw* and the deubiquitinating enzyme *fat facets*¹⁸.

To explain the molecular mechanisms by which RPM-1 regulates synapse development, we searched for genes that function in the same pathway. *rpm-1* mutations enhance the locomotion defects of *syd-2(ju37)*, a mutant with mild defects in active-zone morphology¹⁹ (M.Z. and Y. Jin, unpublished observations). Genes that function in the *rpm-1* pathway might also enhance *syd-2*. In a genetic screen for mutations that enhance the locomotion defects of *syd-2*, we recovered *fsn-1*, a mutant with synaptic defects similar to *rpm-1* mutants.

Both *rpm-1*-null and *fsn-1*-null animals have slightly reduced body length and normal locomotion. Examination of the GABAergic NMJs with the use of a synaptic vesicle marker, green fluorescent protein (GFP)-conjugated synaptobrevin^{19,20} (*juIs1*, Fig. 1c), in *fsn-1* animals revealed that some vesicle pools are 2–5-fold larger than normal (clusters), whereas some regions lack vesicle

pools (gaps) (Fig. 1d). The pan-neural vesicle marker synaptogyrin-GFP (*jsIs219*)²⁰ revealed similar abnormalities in other synapses in *fsn-1* mutants (Fig. 1a, b). In wild-type animals, a single SYD-2 active-zone punctum is associated with the vesicle pool in most GABAergic NMJs (Fig. 1m). In *fsn-1* animals, the large vesicle clusters contain multiple active-zone puncta (Fig. 1n), indicating that they might be ‘overdifferentiated’ synapses. SYD-2 protein is absent from the gap areas, confirming that the gaps are devoid of synapses (not shown). The same phenotype was observed in *rpm-1* mutants⁹ (Fig. 1o). A postsynaptic GABA receptor marker (UNC-49-GFP)⁹ was also clustered in some and missing from other regions of *fsn-1* mutants (Fig. 1j–l). Therefore *fsn-1* synaptic defects include both overdevelopment and underdevelopment of presynaptic and postsynaptic termini, characteristic of the phenotype of *rpm-1* animals.

Consistent with *fsn-1* functioning in the same pathway as *rpm-1* is the observation that *fsn-1;rpm-1* double-null animals display phenotypes similar to that of the *rpm-1* single mutant (Fig. 1d–f). In comparison with the total number of synapses made by the DD motoneurons in wild-type animals (Fig. 1p), *fsn-1*-null animals display a 30% reduction whereas *rpm-1*-null and *rpm-1;fsn-1* double-null animals show a 40% reduction. In addition, *rpm-1* and *fsn-1* mutations synergistically enhance *syd-2* (ref. 19). Both *rpm-1;syd-2* and *fsn-1;syd-2* mutants display a drastic loss of synaptic vesicles in GABAergic NMJs (Fig. 1g–i). Therefore *fsn-1* regulates synaptic development through the *rpm-1* pathway, whereas *rpm-1* has similar but more severe phenotypes in DD motoneurons and in other synapses such as VD and sensory neurons (not shown).

We cloned the *fsn-1* gene by identifying a single open reading frame, C26E6.5, that fully rescues the synaptic defects of *fsn-1* mutants (Supplementary Fig. 1). *fsn-1* encodes a single *C. elegans*

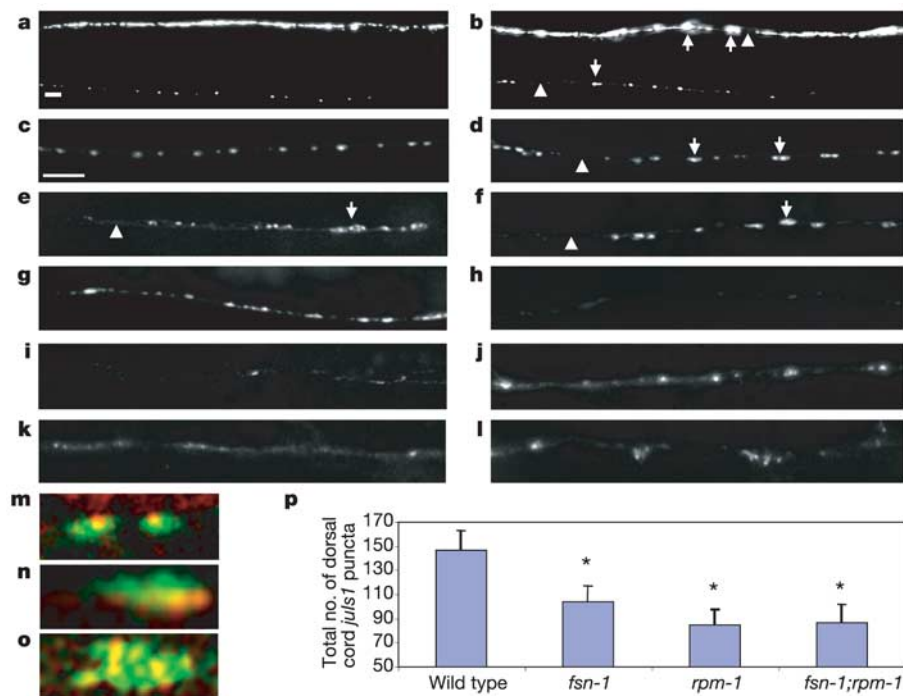


Figure 1 *fsn-1(hp1)* mutants have similar synaptic defects to *rpm-1* animals.

a, b, *jsIs219*, a pan-neural vesicle marker in wild-type (**a**) and *fsn-1* (**b**) animals. *fsn-1* animals have elongated puncta/bulges (arrow) and larger gaps (arrowhead). **c, d**, Vesicle marker *juIs1* in GABAergic motoneurons in wild-type (**c**) and *fsn-1* (**d**) animals. Arrows indicate clustered puncta; arrowheads indicate gaps. **e, f**, Similar *juIs1* phenotypes in *fsn-1* (**d**), *rpm-1* (**e**) and *fsn-1;rpm-1* (**f**) animals. **g–i**, *juIs1* phenotype in *syd-2* (**g**), *fsn-1;syd-2* (**h**) and *rpm-1;syd-2* (**i**) animals. **j–l**, GABA receptor marker *oxIs22* in wild-type (**j**) and

fsn-1 animals that show diffusion (**k**) and clustering (**l**) in different regions. **m**, Double labelling of active zone (SYD-2, red) and vesicles (synaptobrevin-GFP, green) in a wild-type GABAergic NMJ. **n, o**, Enlarged NMJs contain multiple SYD-2 puncta in *fsn-1* (**n**) and *rpm-1* (**o**) mutants. **p**, Quantification of the total number of *juIs1* puncta by DD motoneurons; $n = 30$. Asterisk, $P < 0.001$ compared with wild type (Kruskal–Wallis test). Error bars show standard deviation. Scale bar, 5 μ m.

protein consisting of an F-box domain and a SPRY (for 'spla and ryanodine receptor') domain. Many F-box proteins mediate substrate recognition for SCF ubiquitin-ligase complexes, which contain the invariant components Skp1, Cul1/3 and a RING-finger protein Rbx1⁴⁻⁷. SPRY domains mediate protein-protein interactions²¹. The *fsn-1(hp1)* mutation creates a stop codon before the SPRY domain, and behaves as a genetic and a protein null allele (Supplementary Fig. 1). FSN-1 is highly similar to the predicted *Drosophila* and mouse proteins CG4643 and Q8K3B.

We examined FSN-1 localization with anti-FSN-1 antibodies. FSN-1 is expressed specifically in the nervous system in the nerve processes (Fig. 2a, b). At the synapse, FSN-1 is excluded from the active zone because it does not localize with active-zone protein UNC-10 (Fig. 2c, d). It is present at some periaxial zones because it resides adjacent to the vesicle region (Fig. 2d) and partly overlaps with RPM-1, which is restricted to periaxial zones (Fig. 2e). FSN-1 is also present at non-synaptic, and probably postsynaptic, domains (not shown). At GABAergic NMJs, FSN-1 is absent from the postsynaptic terminal, where GABA receptors localize (Fig. 2f).

Expression of FSN-1 in GABAergic NMJs suggests that it is required in presynaptic neurons. When we expressed FSN-1 cDNA in either GABAergic motoneurons or body wall muscles in *fsn-1*-null animals, only neuronal-FSN-1 expression rescued the defects in GABAergic NMJs. In mosaic animals we found that *fsn-1* is required solely in the neuronal lineage for normal GABAergic NMJs (Supplementary Fig. 2). Therefore FSN-1 is both necessary and sufficient in presynaptic neurons to regulate synapse development in GABAergic NMJs.

Genetic and localization analyses of FSN-1 and RPM-1, and also their predicted structural properties, indicate that they might be components of a neuron-specific E3 complex. We tested this hypothesis biochemically. Because anti-RPM-1 antibody does not work on western analyses, we used an anti-GFP antibody to immunoprecipitate proteins associated with RPM-1 in a transgenic line (*juls58*)⁹ that expresses functional RPM-1-GFP. FSN-1 associated specifically with RPM-1-GFP in *juls58* animals. We also detected CUL-1 and SKR-1, the *C. elegans* homologues of SCF components Cullin and Skp1 (refs 22, 23) (Fig. 2g, left). Using anti-FSN-1 antibody, we again precipitated a complex containing CUL-1 and SKR-1 from both wild-type and RPM-1-GFP animals (Fig. 2g, right). The precipitation of CUL-1 and SKR-1 is dependent on the presence of FSN-1 and RPM-1 (Fig. 2g), indicating that these four proteins might belong to the same complex. In a yeast two-hybrid system C26E6.5(FSN-1) binds SKR-1, whereas SKR-1 associates with CUL-1 (refs 22, 23). Taken together, these results suggest that a previously unknown SCF protein complex, containing FSN-1, RPM-1, CUL-1 and SKR-1, regulates synapse development at periaxial zones.

The FSN-1 complex is unique in that it employs an unusual RING finger subunit, RPM-1. We have determined that Rbx1, the RING finger component in conventional SCF complexes, is unlikely to be present in this complex (Supplementary Fig. 4). Furthermore, the FSN-1 complex depends on RPM-1. RPM-1 is large and consists of several uncharacterized motifs^{8,9}. These regions might be used to modulate or extend the function of currently defined SCF complexes.

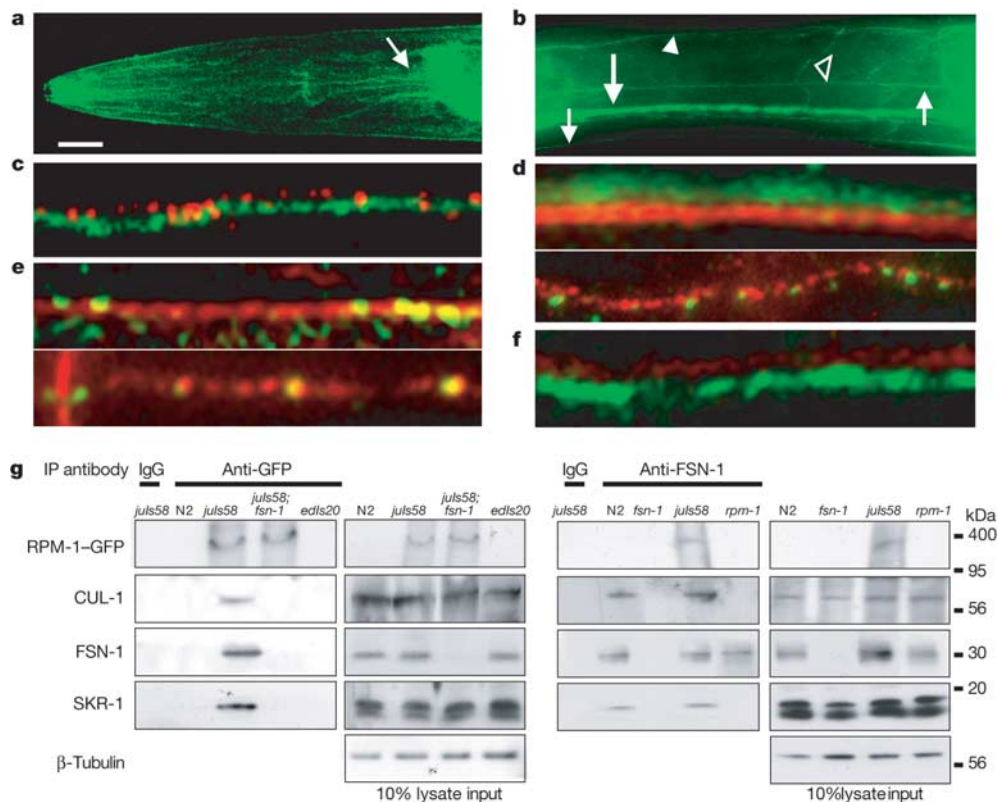


Figure 2 FSN-1 is present at some periaxial zones and non-synaptic regions, and forms a complex with RPM-1. **a, b**, FSN-1 is expressed throughout the nerve system. **a**, Nerve processes in the head (arrow); **b**, processes in the ventral (long arrow), sub-ventral (short arrow), dorsal (arrowhead) nerve cords and circumferential processes (open arrowhead); **c-f**, Immunofluorescent staining for FSN-1 and synaptic components: FSN-1 (green) and active zone (UNC-10, red) (**c**); FSN-1 (red) and vesicles (SNT-1, green) (**d**). Top, the dorsal nerve cord; bottom, the lateral nerve cord where only one layer of synapse is present; FSN-1 (red) and periaxial zone protein RPM-1 (green) in the dorsal nerve cord (top) or

RPM-1-GFP in the lateral nerve cord (bottom) (**e**); FSN-1 (red) and GABA receptor UNC-49-GFP (green) (**f**). Scale bar, 10 μ m (**a, b**), 5 μ m (**c, f**), 0.5 μ m (**d, e**). **g**, FSN-1 forms a complex with RPM-1, CUL-1 and SKR-1. Left, immunoprecipitation (IP) with anti-GFP antibodies from N2 (wild-type), *juls58* (RPM-1-GFP)⁹, *fsn-1(hp1)*; *juls58* and *eds20* (pan-neuronal GFP expression line) lysates. Right, immunoprecipitation with anti-FSN-1 antibody from N2, *juls58*, *fsn-1*-null and *rpm-1*-null lysates. The associated proteins were examined with the indicated antibodies.

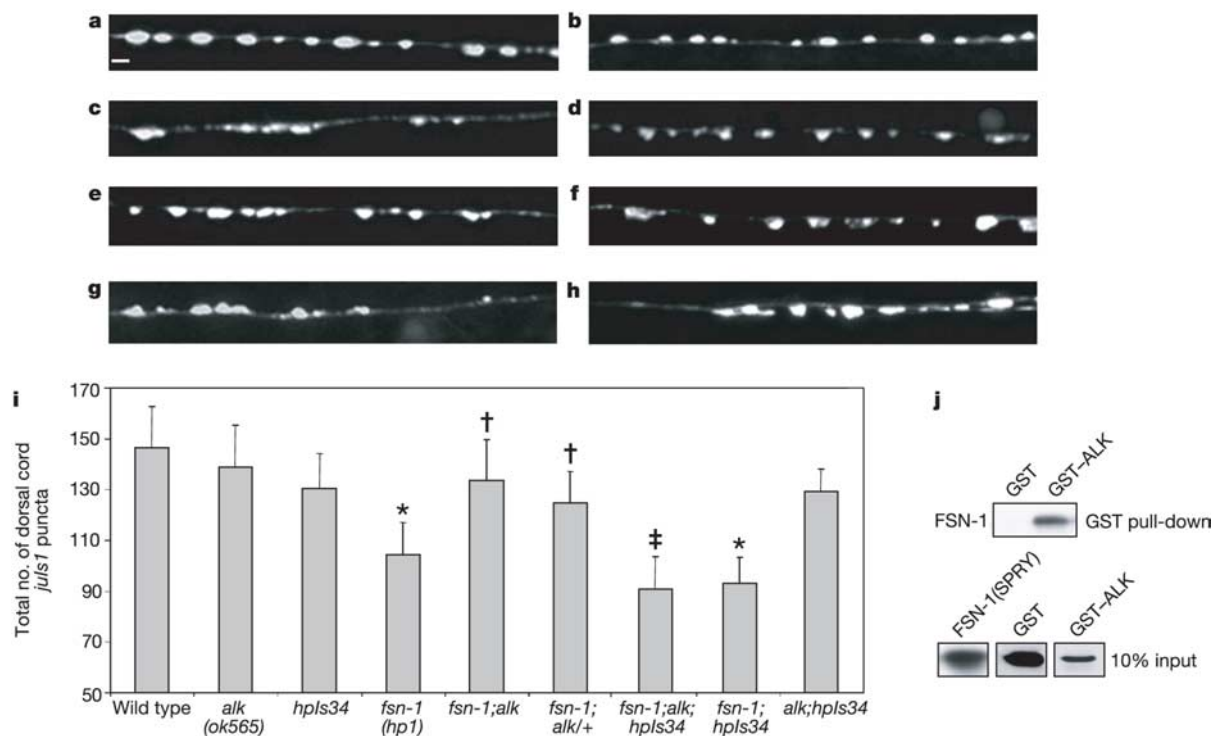


Figure 3 FSN-1 antagonizes the activity of T10H9.2/ALK. **a–h**, Morphology of *juls1* marker in wild-type (**a**), *alk(ok565)* (**b**), *fsn-1(hp1)* (**c**), *fsn-1(hp1);alk(ok565)* (**d**), *fsn-1(hp1);ok565/+* (**e**), *hpls34* (**f**), *fsn-1(hp1);hpls34* (**g**) and *fsn-1(hp1);alk(ok565);hpls34* (**h**) animals. *hpls34* is an integrated ALK–GFP array. Scale bar,

1 μ m. **i**, Quantification of total synapse number by DDs; $n = 30$. Asterisk, $P < 0.001$ compared with wild-type; dagger, $P < 0.001$ compared with *fsn-1*; double dagger, $P < 0.001$ compared with *fsn-1;ok565*– (Kruskal–Wallis test). **j**, GST–ALK/T10H9.2 fusion protein associates with the SPRY domain of FSN-1.

To identify FSN-1 targets, we searched for proteins that physically interact with it. We identified T10H9.2, a *C. elegans* receptor tyrosine kinase homologous to a mammalian proto-oncogene anaplastic lymphoma kinase (ALK)¹⁰, as a binding partner of FSN-1. In a precipitation experiment *in vitro*, T10H9.2/ALK intracellular domain binds to the SPRY domain of FSN-1 (Fig. 3j). If FSN-1 regulates synapse development by downregulating a target, the synaptic defects in *fsn-1* mutants might be restored by a simultaneous decrease in the target's activity. *ok565*, a deletion that truncates the entire transmembrane and intracellular domains of ALK protein, causes little abnormality in synapse morphology (Fig. 3a, b), whereas it suppresses the synaptic defects of *fsn-1*. *fsn-1;ok565/+* animals show some improvement in vesicle size and number (Fig. 3c, e), whereas *fsn-1;ok565* animals display restored synapse number and reduced size of abnormal clusters (Fig. 3c, d). Rescue is variable among individuals and is more robust in larvae than in adults, suggesting that ALK is not the only FSN-1 target. To examine the genetic interactions further, we introduced low-copy-number transgenes expressing ALK, or ALK–GFP under its own promoter (*hpls34*), into *fsn-1;ok565* animals. Although neither array caused significant changes in synapse number in the wild-type (Fig. 3f, i), *fsn-1* (Fig. 3g, i) or *ok565* (Fig. 3i) background, each restored *fsn-1*-like synapse defects in *fsn-1;ok565* animals (Fig. 3h, i). We failed to detect suppression with other *C. elegans* RTK mutants, including *Eph/vab-1*, *egl-15*, *let-23* and *cam-1* (not shown), further supporting the specificity of *ok565*'s effects.

The functional ALK–GFP transgene was used to determine the localization and level of ALK in *fsn-1* and *rpm-1* mutants. In a wild-type background, we detected punctate, low-level expression of ALK–GFP in the nervous system. In head neurons, dim ALK–GFP puncta (red) were associated with active zones (green) (Fig. 4a). In *fsn-1*-null mutants, strong expression of ALK–GFP was present in anterior regions unassociated with active-zone protein (Fig. 4b). In

rpm-1(ju44)-null mutants, ectopic ALK–GFP staining was also detected, although the level was lower than in *fsn-1* (Fig. 4c). Similarly, in *fsn-1* and *rpm-1* nerve processes in the body we detected more dense and intense ALK–GFP puncta than in wild-type animals (Fig. 4d–f). Finally, we showed by western analysis that protein levels of ALK–GFP were elevated in both *fsn-1* (3–4-fold) and *rpm-1* (2–3-fold) mutants, whereas the concentration of another synaptic protein SYD-2 did not increase (Fig. 4g). Therefore the amount and localization of ALK are negatively regulated by FSN-1 and RPM-1. These data suggest that ALK could be a target, or a downstream effector, for FSN-1 in regulating synapse development.

A decrease in *C. elegans* ALK robustly rescues the synaptic morphology defects in *fsn-1* mutants, but only partly rescues *rpm-1* (not shown). It might be caused by the specific *rpm-1* allele. Alternatively, some ALK might be regulated independently through RPM-1. RPM-1 might also function through its other motifs, and this is consistent with a higher degree of synaptic defects in *rpm-1*-null animals.

The t(2;5) translocation in human creates an ectopically expressed constitutively active ALK in lymphocytes, which causes non-Hodgkin's lymphoma¹⁰. Mammalian ALK is expressed in the developing nervous system, with unknown physiological functions²⁴. *Drosophila* ALK is activated by its ligand Jelly belly to specify visceral mesoderm differentiation^{25,26}; however, its function in the nervous system remains elusive. We propose that *C. elegans* ALK might inhibit or destabilize synapse differentiation and that this activity is antagonized by FSN-1, directly or indirectly to stabilize synapse development.

Ubiquitination influences the assembly, disassembly and efficacy of synapses^{27–30}. We demonstrate the existence of a new SCF-like complex comprising an F-box protein FSN-1, the large RING finger protein RPM-1 and the core SCF subunits CUL-1 and SKR-1 in the

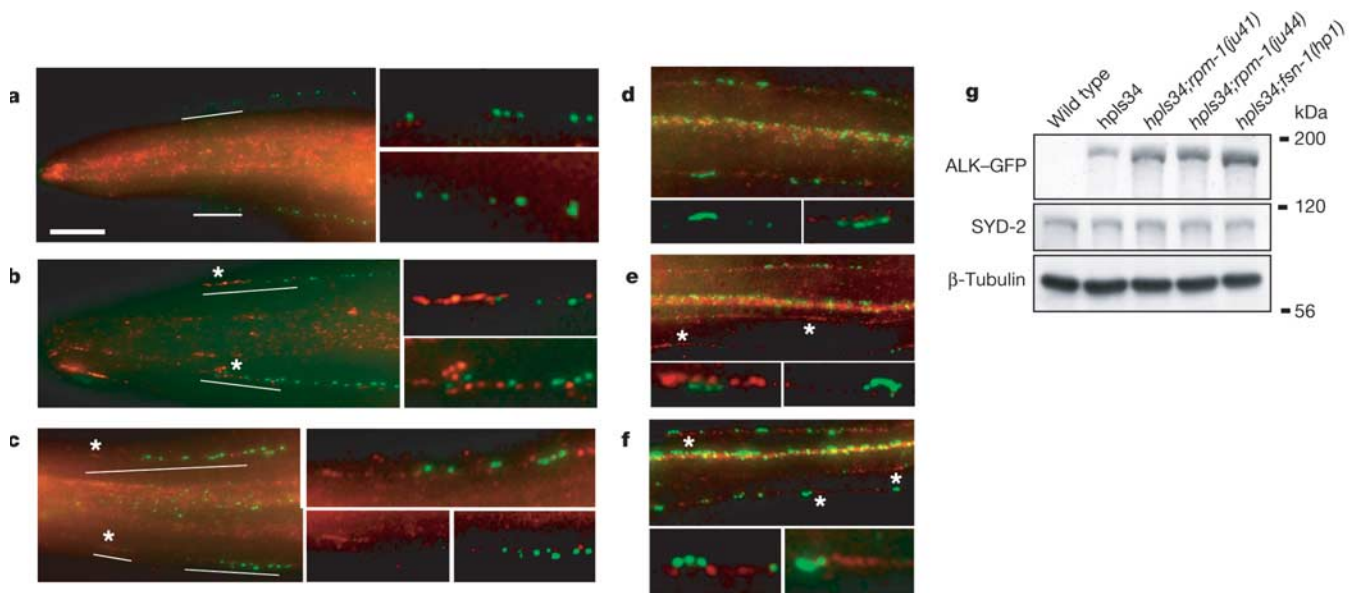


Figure 4 Ectopic and elevated expression of ALK-GFP in *fsn-1* and *rpm-1* mutants. **a–f**, *hpls34* (**a**, **d**), *hpls34; fsn-1(hp1)* (**b**, **e**) and *hpls34; rpm-1(ju44)* (**c**, **f**) animals stained with antibodies against GFP (red) and active zone (UNC-10, green). Enlargements of indicated regions in head (**a–c**) and body (**d–f**) are shown at the right or below the main

panel. Asterisks indicate examples of ectopic and more intense ALK-GFP staining. We overexposed images from the brightest *hpls34* samples. Scale bar, 10 μ m. **g**, Western blot analysis on lysates from synchronized larval cultures.

periaxial zone of *C. elegans* nervous system. This complex imposes localized regulation of presynaptic differentiation by stabilizing synapse position and extent. It does this in part by downregulating the receptor protein tyrosine kinase ALK. The conserved nature of SCF^{FSN-1} components and its potential effector or target suggests that this machinery operates in all metazoans. □

Methods

Strains and genetics

fsn-1(hp1) was isolated in an enhancer screen from EMS-mutagenized CZA65 *syd-2(ju37)*¹⁹ animals. Paralyzed animals were picked from the F₂ progeny. *fsn-1(hp1)* mutation was then outcrossed from the *syd-2* mutation. It was mapped between *mab-21* and *dpy-17* of chromosome III (Supplementary Fig. 1). *hp1* is both a genetic and protein null allele of *fsn-1* (Supplementary Fig. 1). *ok565* was generated by the *C. elegans* Gene Knockout Consortium and backcrossed three times. *hpls34* line was generated by the simultaneous injection and integration of pJH254 (2 ng μ l⁻¹) and RF4 marker (20 ng μ l⁻¹) into *juIs1* animals.

Cloning and constructs of *fsn-1*

Cosmids spanning the *fsn-1* mapping region were injected into *fsn-1(hp1);juIs1* animals and scored for rescuing activity. The following constructs were generated from the rescuing cosmid C26E6 and further tested for their rescuing activity: pJH55 removes the 15-kilobase *NheI* fragment from C26E6, and it was digested with *SnaBI*, *PmlI* or *AgeI* to generate pJH62, pJH63 and pJH125, respectively. pJH122 was generated by inserting the 17-kilobase *AflII/KpnI* fragment from pJH63 into pSL1180 (Pharmacia). pJH252, the genomic clone for T10H9.2, removes the *NcoI* fragments from F44C4. A GFP fragment was inserted into its *KpnI* site to create pJH252.

Antibodies and immunofluorescent staining

His₆-tagged full-length or partial (amino acids 85–332) FSN-1 protein were expressed and purified in bacteria. Chicken antibodies were generated against full-length protein (Covance Immunology Service) and affinity-purified against the partial FSN-1. Whole-mount immunofluorescent staining was performed as described except that the reduction was done in 25 mM sodium borate pH 9.5 (ref. 19). Primary antibodies were used at the following concentration/dilution: 30 μ g ml⁻¹ for anti-FSN-1 antibody, 100 μ g ml⁻¹ for chicken anti-GFP antibody (Chemicon), 5 μ g ml⁻¹ for rabbit anti-GFP antibodies (Molecular Probes), 1:250 dilution for rat RPM-1 antibody, 1:2000 dilution for rabbit SNT-1 and 1:10,000 dilution for rabbit UNC-10 antibodies (M. Nonet, Washington University, St Louis).

Co-immunoprecipitation, GST precipitation and western blotting

C. elegans whole lysate was prepared from 0.5 g frozen pellet by sonication in 2 ml lysis buffer (50 mM Hepes pH 7.5, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM MgCl₂, 1 mM EGTA, 1 mM dithiothreitol and protease inhibitors). For immunoprecipitation experiments, 250 μ g lysate was incubated with 2 μ l rabbit GFP antibody (Clontech) or 300 μ g ml⁻¹ anti-FSN-1 antibody (with 5 μ g rabbit anti-chicken

IgG as the bridging antibody). Protein A beads (20 μ l; Roche) were added and washed with lysis buffer before elution.

For GST precipitation experiments, GST fusion proteins corresponding to T10H9.2 (amino acids 964–1414) were expressed and purified from bacteria. Purified GST fusion protein (10 μ g) was pre-bound to glutathione-Sepharose 4B beads and blocked with 1% BSA and 1% milk. It was incubated with 200 μ g *C. elegans* lysates or 10 μ g purified FSN-1 (amino acids 85–332) protein in binding buffer (20 mM Hepes pH 7.4, 100 mM NaCl, 5 mM dithiothreitol). The beads were washed with binding buffer before elution.

To compare the protein levels of ALK-GFP, synchronized *hpls34*, *hpls34;fsn-1(hp1)*, *hpls34;rpm-1(ju44)* and *hpls34;rpm-1(ju41)* animals were cultured in liquid and harvested at the same development stages. Equal amounts of total protein from each lysate were used for western analysis. This experiment was repeated five times with five sets of independently prepared culture/lysates. Relative ALK-GFP concentrations were quantified with NIH Image. Final antibody concentrations for western blotting were as follows: anti-FSN-1 (0.9 μ g ml⁻¹), rabbit anti-GFP (0.1 μ g ml⁻¹; Chemicon), Cull1 (1:1000) and Skp1 (1:1000) (Y. Xiong, University of North Carolina, Chapel Hill).

Fluorescence microscopy and confocal imaging

Images of GFP markers in live *C. elegans* were captured with the Improvion Openlab system (Quorum Technologies Inc.). Confocal images of antibody stainings were obtained with the Zeiss LSM510 system (Zoology Department, University of Toronto).

Received 13 February; accepted 13 May 2004; doi:10.1038/nature02647.

Published online 20 June 2004.

- Sanes, J. R. & Lichtman, J. W. Development of the vertebrate neuromuscular junction. *Annu. Rev. Neurosci.* **22**, 389–442 (1999).
- Packard, M., Mathew, D. & Budnik, V. Wnts and TGF beta in synaptogenesis: old friends signaling at new places. *Nature Rev. Neurosci.* **4**, 113–120 (2003).
- Scheiffele, P. Cell–cell signaling during synapse formation in the CNS. *Annu. Rev. Neurosci.* **26**, 485–508 (2003).
- Skowron, D. et al. Reconstitution of G1 cyclin ubiquitination with complexes containing SCF^{Grr1} and Rbx1. *Science* **284**, 662–665 (1999).
- Tyers, M. & Jorgensen, P. Proteolysis and the cell cycle: with this RING I do thee destroy. *Curr. Opin. Genet. Dev.* **10**, 54–64 (2000).
- Deshaies, R. J. SCF and Cullin/Ring H2-based ubiquitin ligases. *Dev. Biol.* **15**, 435–467 (1999).
- Ko, H. W., Jiang, J. & Edery, I. Role for Slimb in the degradation of Drosophila Period protein phosphorylated by Doubletime. *Nature* **420**, 673–678 (2002).
- Schaefer, A. M., Hadwinger, G. D. & Nonet, M. L. *rpm-1*, a conserved neuronal gene that regulates targeting and synaptogenesis in *C. elegans*. *Neuron* **26**, 345–356 (2000).
- Zhen, M., Huang, X., Bamber, B. & Jin, Y. Regulation of presynaptic terminal organization by *C. elegans* RPM-1, a putative guanine nucleotide exchanger with a RING-H2 finger domain. *Neuron* **26**, 331–343 (2000).
- Morris, S. W. et al. ALK, the chromosome 2 gene locus altered by the t(2;5) in non-Hodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK). *Oncogene* **14**, 2175–2188 (1997).
- Jin, Y. Synaptogenesis: insights from worm and fly. *Curr. Opin. Neurobiol.* **12**, 71–79 (2002).
- Broadie, K. S. & Richmond, R. E. Establishing and sculpting the synapse in *Drosophila* and *C. elegans*. *Curr. Opin. Neurobiol.* **12**, 491–498 (2002).

13. Sone, M. *et al.* Synaptic development is controlled in the periaxial zones of *Drosophila* synapses. *Development* **127**, 4157–4168 (2000).
14. Wan, H. *et al.* Highwire regulates synaptic growth in *Drosophila*. *Neuron* **26**, 313–329 (2000).
15. McCabe, B. D. *et al.* Highwire regulates presynaptic BMP signaling essential for synaptic growth. *Neuron* **41**, 891–905 (2004).
16. Guo, Q., Xie, J., Dang, C. V., Liu, E. T. & Bishop, J. M. Identification of a large Myc-binding protein that contains RCC1-like repeats. *Proc. Natl Acad. Sci. USA* **95**, 9172–9177 (1998).
17. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70**, 503–533 (2001).
18. DiAntonio, A. *et al.* Ubiquitination-dependent mechanisms regulate synaptic growth and function. *Nature* **412**, 449–452 (2001).
19. Zhen, M. & Jin, Y. The liprin protein SYD-2 regulates the differentiation of presynaptic termini in *C. elegans*. *Nature* **401**, 371–375 (1999).
20. Nonet, M. L. Visualization of synaptic specializations in live *C. elegans* with synaptic vesicle protein-GFP fusions. *J. Neurosci. Methods* **89**, 33–40 (1999).
21. Wang, D., Li, Z., Messing, E. M. & Wu, G. Activation of Ras/Erk pathway by a novel MET-interacting protein RanBPM. *J. Biol. Chem.* **277**, 36216–36222 (2002).
22. Nayak, S. *et al.* The *Caenorhabditis elegans* Skp1-related gene family: diverse functions in cell proliferation, morphogenesis, and meiosis. *Curr. Biol.* **12**, 277–287 (2002).
23. Yamanaka, A. *et al.* Multiple Skp1-related proteins in *Caenorhabditis elegans*: diverse patterns of interaction with Cullins and F-box proteins. *Curr. Biol.* **12**, 267–275 (2002).
24. Iwahara, T. *et al.* Molecular characterization of ALK, a receptor tyrosine kinase expressed specifically in the nervous system. *Oncogene* **14**, 439–449 (1997).
25. Lee, H. H., Norris, A., Weiss, J. B. & Frash, M. Jelly belly protein activates the receptor tyrosine kinase Alk to specify visceral muscle pioneers. *Nature* **425**, 507–512 (2003).
26. Englund, C. *et al.* Jeb signals through the Alk receptor tyrosine kinase to drive visceral muscle fusion. *Nature* **425**, 512–516 (2003).
27. Speese, S. D., Trotta, N., Rodesch, C. K., Aravamudan, B. & Brodie, K. The ubiquitin proteasome system acutely regulates presynaptic protein turnover and synaptic efficacy. *Curr. Biol.* **13**, 899–910 (2003).
28. Hedge, A. N. & DiAntonio, A. Ubiquitin and the synapse. *Nature Rev. Neurosci.* **3**, 854–861 (2002).
29. Colledge, M. *et al.* Ubiquitination regulates PSD-95 degradation and AMPA receptor surface expression. *Neuron* **40**, 595–607 (2003).
30. Pak, D. T. & Sheng, M. Targeted protein degradation and synapse remodeling by an inducible protein kinase. *Science* **302**, 1368–1373 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the *Caenorhabditis* Genetics Centre, Y. Jin, M. Nonet, B. Bamber, J. Kaplan, I. Greenwald, K.-L. Chow, J. Miwa and D. Pilgrim for *C. elegans* strains; The Gene Knockout Consortium for the *T10H9.2(ok565)* deletion strain; Y. Kohara for EST cDNAs; A. Fire for GFP vectors; the Sanger Institute for cosmid; M. Nonet, M. Tyers and Y. Xiong for antibodies; M. Ailion, Y. Jin, B. Meyer, D. Reiner and J. Thomas for communicating unpublished results; Y. Wang for technical support; members of the Zhen laboratory for discussions; and S. Cordes, J. Culotti, Y. Jin, A. Pawson, J. Rossant, M. Tyers and E. Yeh for comments on the manuscript. M.Z. isolated the *fsn-1(hp1)* mutant when she was a postdoctoral fellow in the laboratory of Y. Jin at UC, Santa Cruz. B.A. is supported by a National Institute of Health grant to Y. Jin. This work was funded by a grant to M.Z. from the Canadian Institute of Health Research, Canada.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.Z. (zhen@mshri.on.ca).

Cell fusion-independent differentiation of neural stem cells to the endothelial lineage

Andrew E. Wurmser¹, Kinichi Nakashima^{1,2}, Robert G. Summers¹, Nicolas Toni¹, Kevin A. D'Amour¹, Dieter C. Lie¹ & Fred H. Gage¹

¹The Salk Institute, Laboratory of Genetics, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

²Department of Cell Fate Modulation, Institute of Molecular Embryology and Genetics, Kumamoto University, 2-2-1 Honjo, Kumamoto 860-0811, Japan

Somatic stem cells have been claimed to possess an unexpectedly broad differentiation potential (referred to here as plasticity) that could be induced by exposing stem cells to the extracellular developmental signals of other lineages in mixed-cell cultures^{1–6}. Recently, this and other experimental evidence supporting the existence of stem-cell plasticity have been refuted because stem

cells have been shown to adopt the functional features of other lineages by means of cell-fusion-mediated acquisition of lineage-specific determinants (chromosomal DNA) rather than by signal-mediated differentiation^{1,2,5,7,8}. In this study we co-cultured mouse neural stem cells (NSCs), which are committed to become neurons and glial cells^{9,10}, with human endothelial cells, which form the lining of blood vessels¹¹. We show that in the presence of endothelial cells six per cent of the NSC population converted to cells that did not express neuronal or glial markers, but instead showed the stable expression of multiple endothelial markers and the capacity to form capillary networks. This was surprising because NSCs and endothelial cells are believed to develop from the ectoderm and mesoderm, respectively. Experiments in which endothelial cells were killed by fixation before co-culture with live NSCs (to prevent cell fusion) and karyotyping analyses, revealed that NSCs had differentiated into endothelial-like cells independently of cell fusion. We conclude that stem-cell plasticity is a true characteristic of NSCs and that the conversion of NSCs to unanticipated cell types can be accomplished without cell fusion.

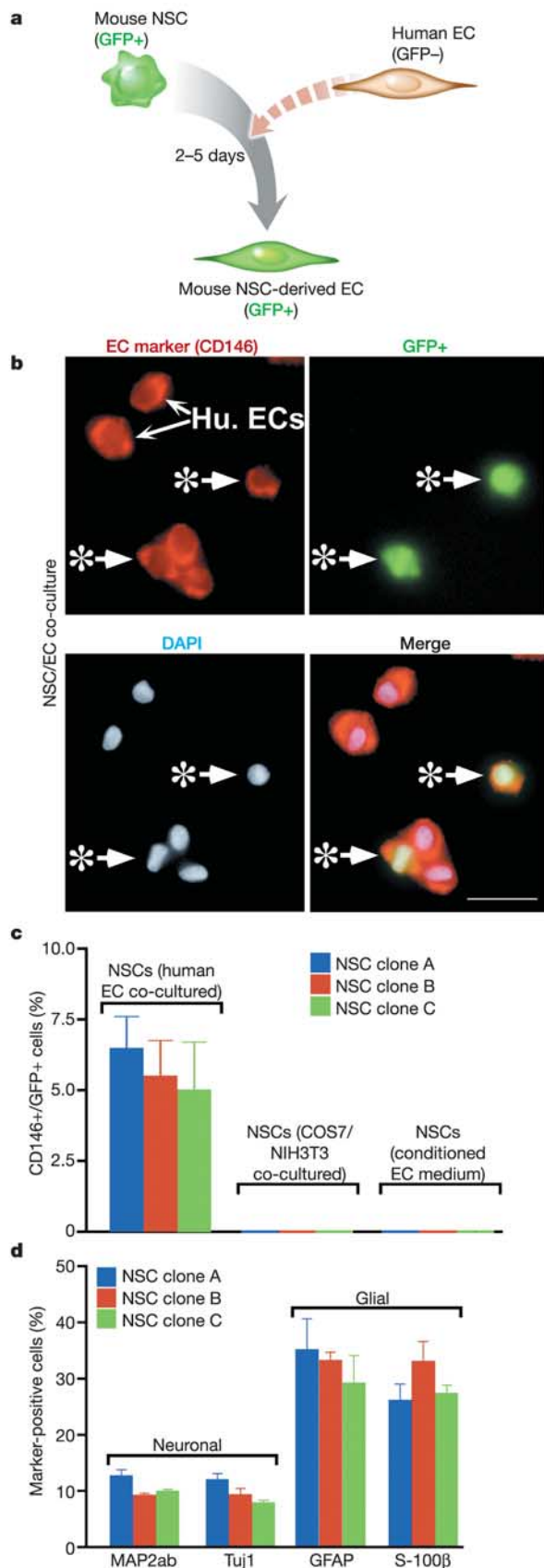
Previous studies have shown that co-cultured stem cells sometimes take on the characteristics of the cell type with which they are cultured^{1–6}. Because NSCs are concentrated at interfaces with the vasculature *in vivo*¹², we reproduced this proximity *in vitro* by co-culturing NSCs with endothelial cells, to determine whether endothelial cells could promote the conversion of NSCs to an endothelial-like cell.

Three clonal lines of NSCs (clones A, B and C), each originating from a single cell, were isolated from mice engineered to constitutively express green fluorescent protein (GFP) in all cell types. Each of our clonally derived, GFP-labelled mouse NSC lines was then co-cultured with a purified population of non-GFP-labelled primary human endothelial cells for 2–5 days (see Methods). All neural and endothelial cells used in these experiments had been passaged only 4–7 times *in vitro*, minimizing the possibility that mutations would occur and affect the outcome of our experiments. To ascertain whether GFP-labelled NSCs could be induced to adopt an endothelial cell phenotype, we probed co-cultures for cells that co-expressed GFP and the endothelial cell adhesion protein, CD146^{13,14} (Fig. 1a).

An antibody specific to CD146 recognized human endothelial cells, but not NSCs, neurons or astrocytes (data not shown). In mouse NSC/human endothelial cell co-cultures, non-GFP-expressing human endothelial cells stained for CD146 in a characteristic cell-surface pattern, with the majority of the CD146 signal concentrated at the cell periphery (Fig. 1b). The presence of human endothelial cells induced approximately 6% of the GFP-labelled cells deriving from NSCs to co-label with CD146 in a staining pattern similar to that of human endothelial cells (Fig. 1b). Whereas multi-nucleated cells have previously been observed after cell fusion^{5,8}, all documented GFP⁺/CD146⁺ cells in our study were uni-nucleate (Fig. 1b).

All of the GFP-labelled NSC clones tested (clones A–C) converted to a CD146-expressing state at a frequency of 5–6.5% after co-culture with human endothelial cells, but not after co-culture with COS7 or NIH3T3 cells under identical conditions (Fig. 1c). The behaviour of these clonal lines was truly representative of the NSC, as a similar percentage (3–9%) of non-clonally purified, low passage (passage 4–7) NSCs expressed CD146 in response to human endothelial cells (data not shown). In the absence of human endothelial cells, mock treatment of NSCs with serum-rich endothelial cell growth medium caused NSC clones A–C to become neurons and astrocytes at frequencies comparable to those previously reported¹⁰ (Fig. 1d). This finding indicates that each of these clonally derived lines behaved as predicted for NSCs and hence that they were free from contamination by non-NSC cell types. Secreted human endothelial cell factors (that is, physical separation of NSCs and human endothelial cells while mixing the respective media)

were also insufficient to bring about GFP⁺/CD146⁺ co-labelled cells, suggesting that induction of CD146 expression requires contact of neural cells and human endothelial cells (Fig. 1c; see below).



Clonal lines of GFP⁺/CD146⁺ cells were then purified. Following 5 days of co-culture with human endothelial cells, 600 GFP-labelled cells derived from NSCs were subjected to a single-cell-per-well, fluorescence-activated cell sort (FACS) for GFP-positive cells. GFP-positive cells were then clonally expanded in serum-rich endothelial cell growth medium and screened for the expression of CD146. Three clonal lines (clones 1, 2 and 3) were established at a frequency of 0.5%; lower than the 6% incidence of GFP⁺/CD146⁺ cells in co-culture. This discrepancy may be due to cell death during the FACS procedure.

Similar to human endothelial cells, 84–88% of cells within each clonal population of NSC-derived endothelial cell candidates expressed CD146, as seen by indirect immunofluorescence, whereas mock-treated NSCs exposed to serum-rich endothelial cell growth medium in the absence of human endothelial cells did not (Fig. 2a, b). Indeed, GFP and CD146 expression were maintained by clones 1–3 of NSC-derived endothelial cell candidates even at late passage (passage 22; data not shown), suggesting that GFP⁺/CD146⁺ clonal cell lines did not require the continued presence of human endothelial cells to maintain expression of this endothelial cell marker. We found that a significant percentage of the human endothelial cell population labelled with the characterized endothelial cell markers vascular endothelial (VE) cadherin (95%), von Willebrand factor (VWF; 69%), isolectin GS-IB₄ (lectin; 80%) and CD31 (85%) (Fig. 2a, b)^{15–19}. The percentage of cells within each GFP⁺/CD146⁺ clonal line that stained positive for these endothelial markers was comparable: VE cadherin (84–88%), VWF (82–92%), lectin (78–80%) and CD31 (77–86%) (Fig. 2a, b). A second VE cadherin antibody also recognized 28% of human endothelial cells and 56–76% of NSC-derived endothelial cell candidates (data not shown). Detergent did not influence the distribution of VE cadherin, confirming the predominantly cell-surface localization of this cell adhesion protein (data not shown)^{15,16}. A high percentage of cells within clonal GFP⁺/CD146⁺ cell lines also expressed the endothelial proteins, endoglin (86–92%), connexin 40 (89–93%), endothelial nitric oxide synthase (80–86%), thrombomodulin (81–91%) and claudin 5 (89–95%; Supplementary Fig. 1a). Additionally, even though each GFP⁺/CD146⁺ clone originated from NSCs, none of these cell lines expressed markers associated with NSCs or the lineages known to develop from NSCs, such as nestin (NSC), GFAP (astrocyte), RIP (oligodendrocyte), or Map2ab/Tuj1 (neuron) (Fig. 2a, b). In the absence of human endothelial cells, NSCs mock-treated with serum-rich endothelial cell medium expressed Map2ab, Tuj1 and GFAP (Fig. 2a, b). NSCs maintained under serum-free neural growth conditions were positive for nestin, but did not label with CD146, VE cadherin, VWF, CD31 or lectin (data not shown).

Reverse transcriptase-polymerase chain reaction confirmed that NSCs that were exposed to neural growth conditions or mock-

Figure 1 NSCs can be induced to express the endothelial marker, CD146. **a**, GFP-labelled mouse NSCs were mixed with non-GFP-labelled human endothelial cells (ECs) and co-cultured under serum-rich endothelial cell growth conditions for 2–5 days. We then screened for NSC-derived endothelial cells, which were predicted to co-express GFP and the endothelial marker, CD146. **b**, Fluorescent images of a single field of co-cultured human endothelial cells and GFP-labelled NSCs. The endothelial cell marker CD146, GFP, nuclear DAPI and a merge of these three images are shown. Human endothelial cells and GFP⁺/CD146⁺ cells (*) are indicated. Scale bar, 30 μ m. **c**, Quantification of GFP⁺/CD146⁺ cells derived from three clonal lines of NSCs (clones A–C) after co-culture with either COS7, NIH3T3 or human endothelial cells, or after treatment of NSCs with human endothelial cell-conditioned medium for 5 days. Mean values $\pm$ s.d. are shown. **d**, Mock treatment of three NSC clones (A–C) with serum-rich endothelial cell medium for 5 days induced differentiation to neuronal (Map2ab or Tuj1 markers) and glial (GFAP or S-100 β markers) lineages. Mean values $\pm$ s.d. are shown.

treated with endothelial cell medium were enriched for transcripts of the neural lineages (nestin or Map2ab/GFAP, respectively). These transcripts were down-regulated in NSC-derived GFP⁺/CD146⁺ cells (Fig. 2c). Instead, GFP⁺/CD146⁺ cells expressed messages for endothelial VE cadherin, CD31, Flk-1, Tie-2, endoglin, connexin 40, endothelial nitric oxide synthase and thrombomodulin^{16,19,20} (Fig. 2c; Supplementary Fig. 1b).

After determining that these three lines of NSC-derived endothelial cell candidates expressed multiple markers of endothelial cells and lacked expression of neural markers, we investigated the possibility that these cells might also have adopted the functional properties of endothelial cells. One ultra-structural feature of endothelial cells expressing VWF is the presence of a unique class of secretory vesicles, known as Weibel–Palade bodies. Electron microscopy revealed that these vesicles, which are 200–400 nm in diameter and characterized by an electron-dense luminal matrix²¹, were present in human endothelial cells and clones 1–3 of NSC-derived endothelial cell candidates (Fig. 3a). Another func-

tional hallmark of differentiated endothelial cells is the capacity to form capillaries. Although this property is not manifested during growth conditions, it can be induced by plating endothelial cells onto a collagen/laminin-coated substratum (Matrigel)²². Therefore, non-human endothelial cell co-cultured NSCs, human endothelial cells and clonally purified populations of NSC-derived endothelial cell marker-positive cells were assayed for Matrigel-dependent vascular cord formation *in vitro*. Time-lapse imaging revealed that each clone of NSC-derived endothelial cell marker-expressing cells and human endothelial cells, initially plated as dissociated cells, began aggregating and elongating within 2 h. Vascular cord formation for each NSC-derived, endothelial-like clone and human endothelial cells proceeded with identical kinetics and was complete by the 12-h time point (Fig. 3b; Supplementary Fig. 2). In contrast, whereas mock-treated NSCs remained motile throughout most of the assay, these cells failed to form the same structures (Fig. 3b). Collectively, this body of evidence suggests that the co-culture of NSCs with human endothelial cells can divert NSC-derived

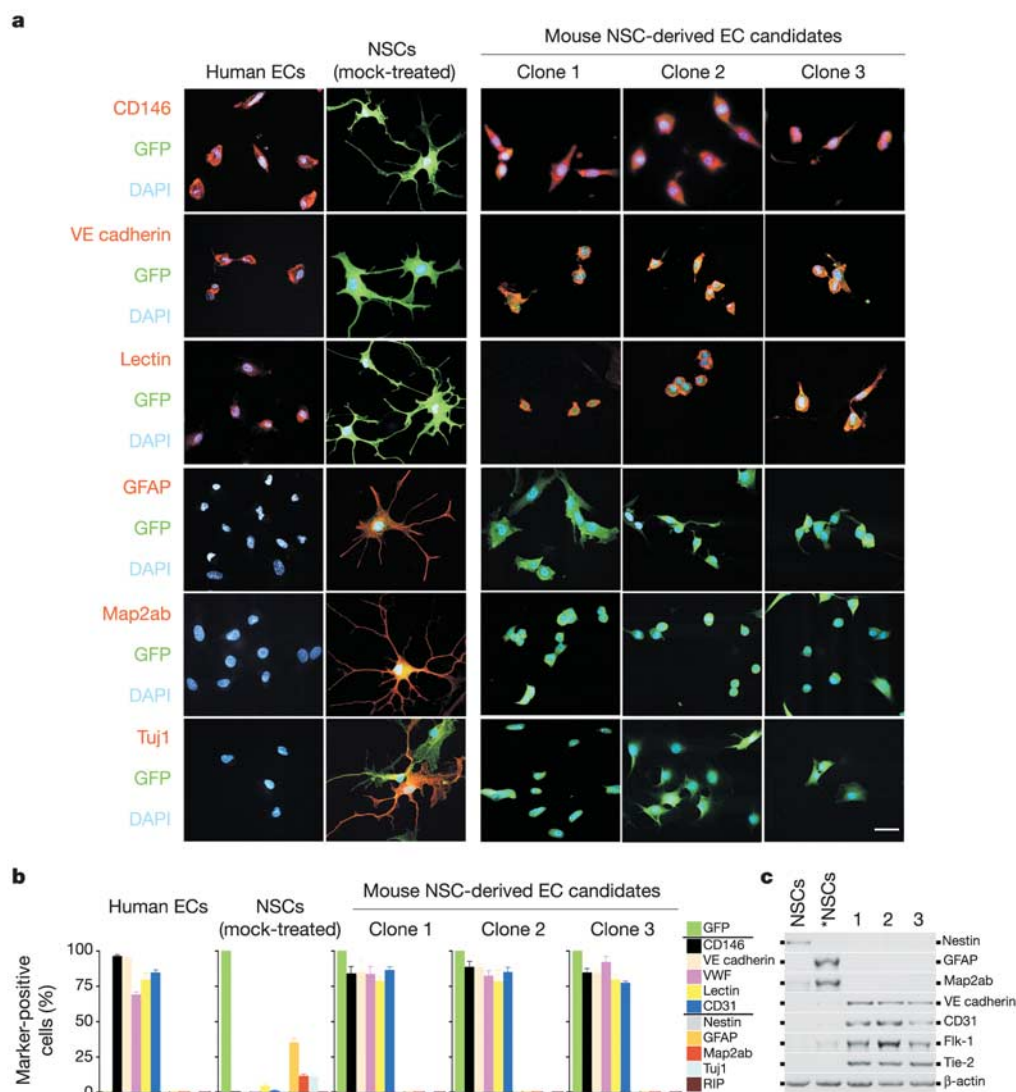


Figure 2 Endothelial markers are expressed by NSC-derived endothelial cell candidates clonally purified from human endothelial cell co-culture. **a–c**, Expression of endothelial, neuronal and glial markers in human endothelial cells, mock-treated NSCs and three clonal lines of NSC-derived endothelial cell candidates (clones 1–3). **a**, Fluorescent images of each cell type. Scale bar, 20 μ m. **b**, Quantification of the percentage of marker-

positive cells by fluorescent imaging. Mean values $\pm$ s.d. are shown. **c**, Reverse transcriptase-polymerase chain reaction analysis of nestin, Map2ab, GFAP, VE cadherin, CD31, Flk-1, Tie-2 and β -actin messenger transcript expression in NSCs maintained under growth conditions, mock-treated NSCs (column marked with an asterisk) and NSC-derived endothelial cell candidates (columns labelled 1–3).

cells from the neural lineages, instead inducing endothelial characteristics.

Having been observed to form liver, epithelial, blood and muscle cells, NSCs were initially proposed to have a highly plastic differentiation programme that could readily bypass the neural lineages^{3,23,24}. A growing number of studies now suggest that stem cells, including NSCs, adopt the phenotypes of other cells by fusion, an event that may previously have been mistaken for stem-cell plasticity^{1,2,5,7,8}. It is currently unclear whether cell fusion can account exclusively for all reports of stem-cell plasticity^{4,8,25}. Nonetheless, the conversion of NSCs to endothelial-like cells might entail the cell-fusion-mediated formation of mouse NSC–human endothelial cell hybrids that epistatically express human endothelial cell phenotypes and function.

Analysing the chromosomal content of cultured stem-cell derivatives has proven to be effective in the detection of stem-cell fusion^{1,2,8}. On the basis of metaphase chromosomal spreads of previously reported mouse–human hybrids, low-passage mouse NSC–human endothelial cell hybrids would be predicted to have approximately 40 mouse and 10–40 human chromosomes^{26,27}. As shown in Fig. 4a, human endothelial cells manifested 46 chromosomes that were metacentric or acrocentric, whereas mouse NSCs contained 40 telocentric chromosomes. Consistent with a non-fusion event, the screening of 141 chromosomal spreads revealed approximately 40 telocentric mouse chromosomes, but never any human chromosomes, within low-passage endothelial-like cells derived from NSCs (Fig. 4a). Additionally, a species-specific antibody that recognizes human ribonuclear proteins reacted strongly with the nuclei of human endothelial cells but did not notably label the nuclei in any of the three clonal, NSC-derived endothelial-like lines (Fig. 4b). These experiments strongly suggest that each of the three mouse NSC-derived endothelial-like cell lines maintained endothelial phenotypes independently of genomic line-

age determinants acquired from human endothelial cells by cell fusion.

To address a potential role for cell fusion at the initiation of the NSC-to-endothelial cell transition, human endothelial cells were killed with paraformaldehyde before co-culture with NSCs. Paraformaldehyde pre-fixation of human endothelial cells is predicted to stabilize cellular components, rendering human endothelial cells refractory to cell fusion (Fig. 4c). This method has been shown to prevent fusion of live sperm cells with fixed sea urchin eggs, while not disrupting receptor-mediated recognition and association of these cell types²⁸. As expected, paraformaldehyde treatment disrupted the exclusion of propidium iodide by human endothelial cells, indicating that these cells were non-viable (Fig. 4d). After extensive washing to remove excess paraformaldehyde, GFP-labelled NSCs were cultured with fixed human endothelial cells for 5 days. During the course of this study, we found that lectin preferentially bound to mouse NSC-derived endothelial-like cells with several-fold higher affinity than human endothelial cells. Lectin was therefore used as an endothelial probe to facilitate identification of NSC-derived endothelial-like cells in the context of fixed human endothelial cell co-cultures. This procedure led to the conversion of 2–4% of NSCs to GFP⁺/lectin⁺ double-labelled cells relative to fixed COS7 cell co-cultured NSCs (Fig. 4e, f). Collectively, these experiments indicate that cell fusion did not play a part in the initial expression or maintenance of endothelial traits by NSC-derived cells. The data instead support the existence of relatively paraformaldehyde-resistant cell surface molecules (such as transmembrane proteins) that convey endothelial differentiation or survival signals from human endothelial cells to NSC-derived cells.

Stem-cell plasticity was thought to form the foundation for one of the multiple prospective uses of adult stem cells in regenerative medicine²⁹. Data demonstrating that stem cells could fuse with and subsequently adopt the phenotypes of other cell types indicated that the very co-culture assays originally interpreted to support plasticity instead were artefacts of cell fusion^{1–6}. By showing that NSCs unexpectedly differentiated to the endothelial lineage in the absence of cell fusion, we reintroduce the idea that NSCs are inherently plastic.

This demonstration of NSC plasticity *in vitro* may reflect the potential for NSCs to differentiate to endothelial cells *in vivo*. Consistent with this possibility, when clonal GFP-labelled NSCs were transplanted into the telencephalon of embryonic day-14 mice, 1.6% of NSCs converted to cells that co-expressed endothelial phenotypes (that is, bound to lectin and expressed VE cadherin) by embryonic day 16 (Fig. 4g, h). Most transplanted NSCs (>80%) differentiated to glial cells (data not shown). This *in vivo* assay system did not recapitulate the high ratio of endothelial cells to NSCs in our *in vitro* experiments, possibly explaining the relative frequency of NSC-to-endothelial cell conversion *in vivo* (1.6%) versus *in vitro* (6%). Nonetheless, NSC-derived endothelial-like cells showed the characteristic cup-shaped morphology of endothelial cells and also labelled for VE cadherin at intercellular junctions with endogenous endothelial cells (Fig. 4g, h). Optical sectioning of NSC-derived endothelial cells by confocal microscopy showed that these cells had a single nucleus, suggesting that NSCs had differentiated to endothelial-like cells *in vivo* (Fig. 4g, h).

This work encourages a new look at angiogenesis, the process whereby endothelial cells dissociate from pre-existing vessels, proliferate and aggregate to form new vessels directed at oxygen- and nutrient-deprived tissue¹¹. The traditional view of angiogenesis suggests that newly formed endothelial cells can arise through the proliferation of pre-existing endothelial cells¹¹. An alternative possibility raised by this work is that NSCs give rise not only to neurons and glial cells but also to endothelial cells that promote vascularization of the surrounding tissue. Further experimentation

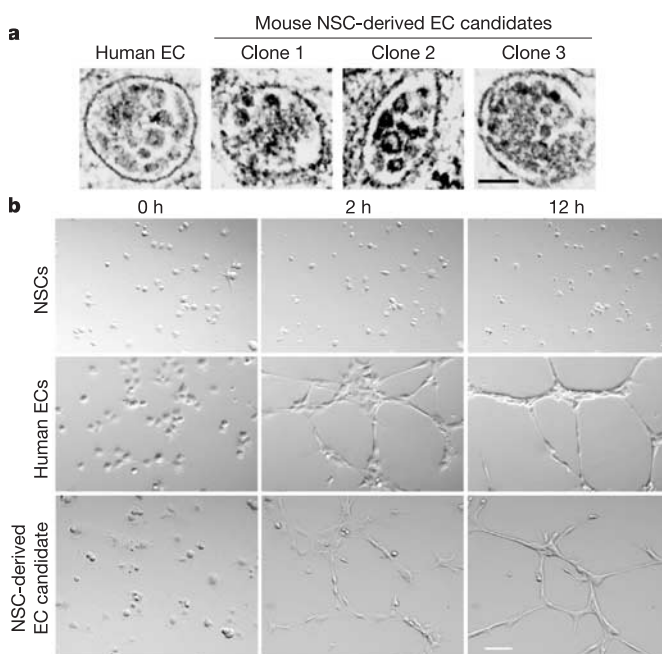


Figure 3 NSC-derived endothelial cell candidates show endothelial Weibel–Palade bodies and form capillary networks *in vitro*. **a**, Electron microscopy of subcellular Weibel–Palade bodies in human endothelial cells and NSC-derived endothelial cell candidates. Scale bar, 100 nm. **b**, NSCs, human endothelial cells and NSC-derived endothelial cell candidates (clone 1 is shown) were introduced onto Matrigel-coated plates at the 0 h time point and monitored for their capacity to form vascular cords. Scale bar, 100 μm.

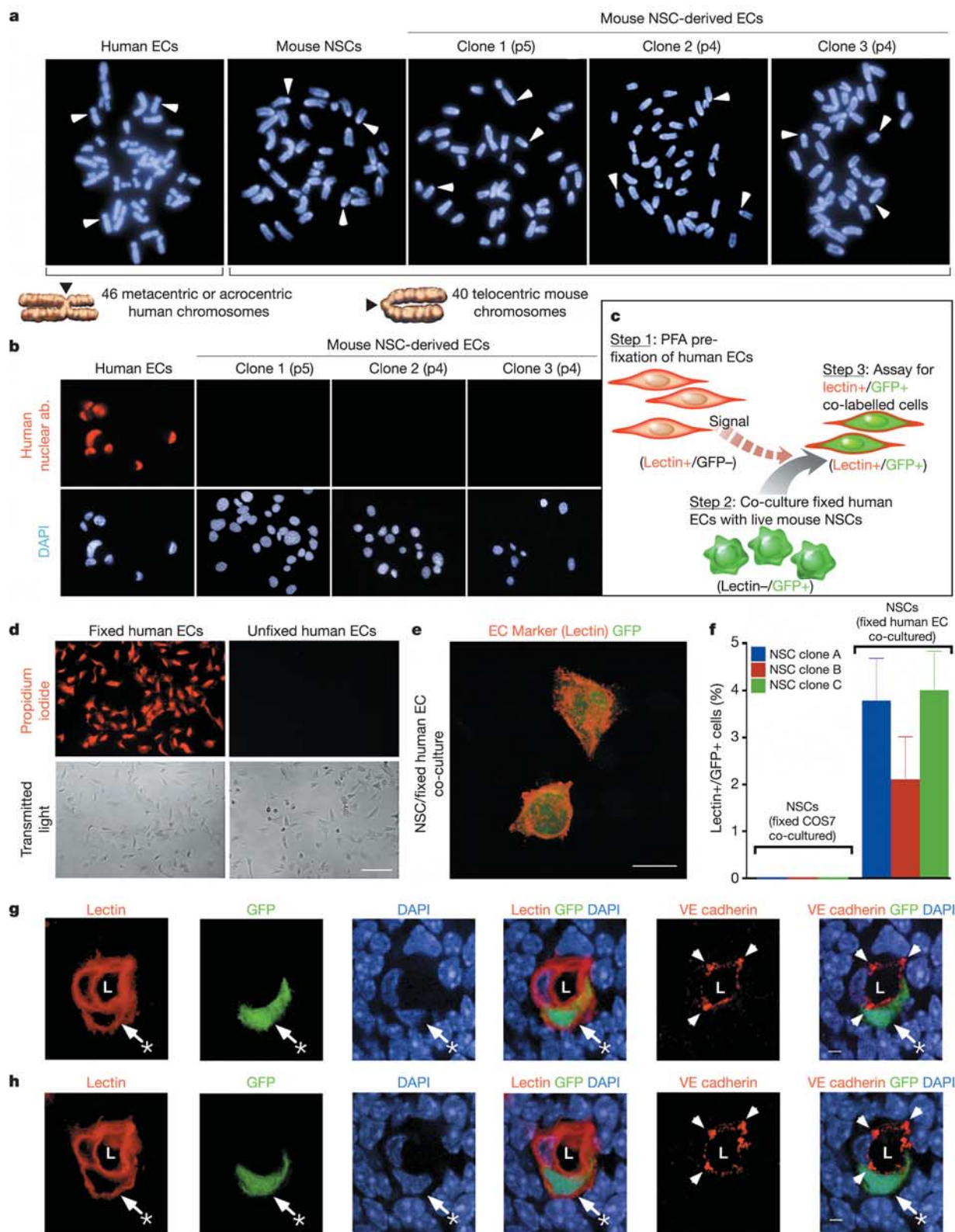


Figure 4 Cell fusion is unlikely to contribute to the generation of NSC-derived endothelial-like cells. **a**, Metaphase chromosome spreads of human endothelial cells, mouse NSCs and clones 1–3 of mouse NSC-derived endothelial-like cells are shown. Clones 1–3 were karyotyped at passages 5 (p5), 4 (p4) and 4, respectively. Several centromeres are highlighted (arrowheads). **b**, Human endothelial cells and clones 1–3 of NSC-derived endothelial-like cells at passages 5, 4 and 4, respectively, stained with human-specific ribonuclear protein antibody and DAPI. **c**, Procedure for the culture of live GFP-labelled NSCs with fixed human endothelial cells. **d**, Fixed human endothelial cells stained with

propidium iodide. Scale bar, 100 μ m. **e**, **f**, Fixed human endothelial cells induced clones A–C of GFP-labelled NSCs to bind the endothelial probe lectin. Scale bar, 20 μ m (**e**). Mean values $\pm$ s.d. are shown (**f**). **g**, **h**, GFP-labelled NSCs converted to lectin and VE cadherin co-expressing endothelial cells within the neocortex of the subventricular zone in embryonic day-16 mice (indicated by arrow and an asterisk). Two sections of one sample capillary are shown. The lumen of the capillary (L) and the adherent VE cadherin junctions between endothelial cells (arrowheads) are highlighted. Scale bar, 10 μ m.

will be necessary to distinguish whether NSCs contribute to vascular outgrowths at a frequency inherent to the normal physiology of the embryonic and adult central nervous systems. □

Methods

Cell culture

Whole brain NSCs of adult mice that ubiquitously express GFP (C57BL/6-TGN (ACTbEGFP)10sb; The Jackson Laboratory) were purified in a two-step process first entailing the fractionation of mitotic NSCs from non-mitotic cells using a Percoll gradient¹⁰. Gradient-enriched NSCs were then clonally isolated by a single-cell-per-well FACS-based selection of GFP-positive cells. On the day of plating, visual verification by phase contrast and GFP-fluorescence microscopy confirmed that clonal lines A–C originated from a single cell. NSCs were proliferated in DMEM/F-12 medium (Omega Scientific) supplemented with N2 (Gibco), 2 mM L-glutamine, 20 ng ml⁻¹ fibroblast growth factor-2 and 20 ng ml⁻¹ epidermal growth factor. Purified human umbilical artery endothelial cells (human endothelial cells) were grown in EGM2–MV endothelial medium that contains vascular endothelial growth factor, fibroblast growth factor-2, epidermal growth factor, insulin growth factor I and 5% serum (BioWhittaker/Clonetics).

Clonal lines of NSCs were grown as a monolayer and shifted from growth medium into serum-rich EGM2–MV medium for 5 days to quantify NSC differentiation to neurons, glial cells and CD146-expressing cells.

For co-cultures, NSCs and human endothelial cells (or COS7/NIH3T3) were trypsinized and resuspended in NSC growth or EGM2–MV medium. Seventy-five NSCs were then mixed with 750–7,500 human endothelial cells in 400 µl EGM2–MV per well of a 24-well plate (Costar). Co-culture was continued until 80% confluence was attained (2–5 days). Cells were fed 200 µl of EGM2–MV on day 3. Cells that stained positive for CD146 in the characteristic staining pattern shown in Fig. 1b were scored as positive. Four thousand human endothelial cells were plated per well of a 24-well plate and grown to 80% confluence before being fixed with 0.5% paraformaldehyde for 15 min at 25 °C. Fixed cells were washed extensively, first with PBS and then overnight with EGM2–MV. The next day, 300 GFP-labelled NSCs (clones A–C) were cultured with fixed human endothelial cells for 5 days and then stained with lectin, as described below. The frequency of conversion of NSCs to GFP⁺/lectin⁺ cells was calculated by normalizing the number of GFP⁺/lectin⁺ cells resulting from human endothelial cell co-culture with the number of GFP⁺/lectin⁺ cells arising from paraformaldehyde-fixed COS7 co-culture.

Capillary formation

Endothelial cells were induced to form vascular cords by pre-coating 24-well plates with collagen/laminin (Matrigel; BD Biosciences). Each cell type was plated in EGM2–MV lacking serum, epidermal growth factor and insulin growth factor I and then was incubated at 37 °C for 12 h. Cells were monitored by Hoffman modulation contrast optics using a × 10 lens.

Microscopy

Culture medium was washed away for 10–15 min with 0.1 M TBS buffer before fixation with 4% paraformaldehyde. Fixed cells were incubated with primary antibodies overnight in 0.1 M TBS buffer supplemented with 3% normal donkey serum and 0.25% Triton X-100, or with lectin for 1 h in 10 mM HEPES buffer pH 7.6, 150 mM NaCl and 1 mM CaCl₂. The primary antibodies used were: CD146 (1 µg ml⁻¹, mouse; Chemicon), Map2ab (1:250, mouse; Sigma), Tuj1 (1:1,000, mouse; Covance Inc.), GFAP (1:2,500, rabbit; Dako), S-100β (1:2,500, rabbit; Sigma), VE cadherin (3 µl ml⁻¹, rabbit; Alexis Biochemicals), anti-mouse VE cadherin (1:100, rat; Research Diagnostics), VWF (1:1,000, sheep; laboratory reagent), CD31 (2.5 µl ml⁻¹, rat; Caltag Laboratories), nestin (1:1,000, mouse; Pharmingen), RIP (1:50, mouse; Hybridoma Bank), human nuclei (1:300, mouse; Chemicon), endothelial nitric oxide synthetase (1.5 µl ml⁻¹, mouse; Research Diagnostics), connexin 40 (12 µl ml⁻¹, rabbit; Zymed), claudin 5 (6 µl ml⁻¹, rabbit; Lab Vision), thrombomodulin (6 µl ml⁻¹, mouse; Lab Vision) and endoglin (12.5 µl ml⁻¹, mouse; Lab Vision). Lectin was used at 2 µg ml⁻¹ (Alexa 594- or 647-conjugated; Molecular Probes). Where applicable, CY3- or Texas Red-conjugated secondary antibodies specific to the appropriate species were used (1:250; Jackson Laboratory).

Images were obtained with either a multi-photon equipped confocal system (Bio-Rad) or an Eclipse E800 microscope (Nikon)/Spot Camera system (Diagnostic Instruments).

Quantification of marker expression was carried out using an Eclipse E800 microscope (Nikon)/Spot Camera system (Diagnostic Instruments). Each marker was quantified as a percentage of 4,6-diamidino-2-phenylindole (DAPI)-positive cells that were also marker positive. Results of three to four experiments are shown.

For electron microscopy, EGM2–MV was removed and cultured cells were fixed by the addition of 3% glutaraldehyde/0.1 M phosphate buffer followed by 0.1% osmium tetroxide. Cells were dehydrated in 80% ethanol, scraped off the plate with a Teflon blade and pelleted. Cells were then embedded in EPON resin mixture. Ultrathin sections of a grey interference colour (~40 nm) were observed with a Jeol 100CX electron microscope.

Karyotyping

Cells were conditioned in EGM2–MV supplemented with 60 µg ml⁻¹ bromodeoxyuridine (Sigma), 0.3 µg ml⁻¹ PuDr (5-fluoro-2'-deoxyuridine; Sigma) and 2 µl ml⁻¹ colcemid (Gibco BRL) for 5 h at 37 °C. After trypsinization, cells were washed with PBS and then incubated in 0.06 M KCl for 15 min at 37 °C. Chromosomes were then fixed in methanol/acetic acid (3:1) overnight, attached to slides by flaming and stained with DAPI.

Reverse transcriptase-polymerase chain reaction

A complementary DNA template was synthesized using SuperScript reverse transcriptase-polymerase chain reaction kit (Invitrogen) and subjected to 20–33 rounds of polymerase chain reaction in the presence of one unit of Taq polymerase (Promega), 4 mM MgCl₂, 400 µM deoxyribonucleotide triphosphates and 1 µg of each oligonucleotide primer per 50 µl reaction. Nested primers internal to the initial primers were used to confirm the identity of first-round polymerase chain reaction fragments. All primers were designed to hybridize with the murine version of each transcript tested. β-actin was used as an internal standard.

In utero transplantation of NSCs

Approximately 1 µl containing 10⁵ GFP-expressing NSCs (line 2, clonally-derived NSCs) were injected into the lateral ventricle of embryonic day-14 ICR mice *in utero*. The NSC-transplanted embryonic brains were collected on embryonic day 16 and fixed with 4% paraformaldehyde overnight. Fixed brains were sectioned at 14 µm thickness on a Microm HM 560 cryostat and probed with lectin (Alexa 647-conjugated; Molecular Probes) and VE cadherin (1:30; Research Diagnostics), as described above. Endothelial marker-expressing cells originating from GFP-labelled NSCs were quantified as a percentage of GFP-labelled cells that integrated into neural tissue.

Received 15 January; accepted 23 April 2004; doi:10.1038/nature02604.

1. Ying, Q. L., Nichols, J., Evans, E. P. & Smith, A. G. Changing potency by spontaneous fusion. *Nature* **416**, 545–548 (2002).
2. Terada, N. *et al.* Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature* **416**, 542–545 (2002).
3. Galli, R. *et al.* Skeletal myogenic potential of human and mouse neural stem cells. *Nature Neurosci.* **3**, 986–991 (2000).
4. Alexanian, A. R. & Sieber-Blum, M. Differentiating adult hippocampal stem cells into neural crest derivatives. *Neuroscience* **118**, 1–5 (2003).
5. Spees, J. L. *et al.* Differentiation, cell fusion, and nuclear fusion during *ex vivo* repair of epithelium by human adult stem cells from bone marrow stroma. *Proc. Natl Acad. Sci. USA* **100**, 2397–2402 (2003).
6. Condorelli, G. *et al.* Cardiomyocytes induce endothelial cells to trans-differentiate into cardiac muscle: implications for myocardium regeneration. *Proc. Natl Acad. Sci. USA* **98**, 10733–10738 (2001).
7. Vassilopoulos, G., Wang, P. R. & Russell, D. W. Transplanted bone marrow regenerates liver by cell fusion. *Nature* **422**, 901–904 (2003).
8. Wang, X. *et al.* Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature* **422**, 897–901 (2003).
9. Turner, D. L. & Cepko, C. L. A common progenitor for neurons and glia persists in rat retina late in development. *Nature* **328**, 131–136 (1987).
10. Luskin, M. B., Pearlman, A. L. & Sanes, J. R. Cell lineage in the cerebral cortex of the mouse studied *in vivo* and *in vitro* with a recombinant retrovirus. *Neuron* **1**, 635–647 (1988).
11. Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671–674 (1997).
12. Palmer, T. D., Willhoite, A. R. & Gage, F. H. Vascular niche for adult hippocampal neurogenesis. *J. Comp. Neurol.* **425**, 479–494 (2000).
13. Bardin, N. *et al.* Identification of CD146 as a component of the endothelial junction involved in the control of cell–cell cohesion. *Blood* **98**, 3677–3684 (2001).
14. St Croix, B. *et al.* Genes expressed in human tumor endothelium. *Science* **289**, 1197–1202 (2000).
15. Lampugnani, M. G. *et al.* A novel endothelial-specific membrane protein is a marker of cell–cell contacts. *J. Cell Biol.* **118**, 1511–1522 (1992).
16. Breier, G. *et al.* Molecular cloning and expression of murine vascular endothelial-cadherin in early stage development of cardiovascular system. *Blood* **87**, 630–641 (1996).
17. Wagner, D. D., Olmsted, J. B. & Marder, V. J. Immunolocalization of von Willebrand protein in Weibel–Palade bodies of human endothelial cells. *J. Cell Biol.* **95**, 355–360 (1982).
18. Goldstein, I. J. & Hayes, C. E. The lectins: carbohydrate-binding proteins of plants and animals. *Adv. Carbohydr. Chem. Biochem.* **35**, 127–340 (1978).
19. Newman, P. J. & Albelda, S. M. Cellular and molecular aspects of PECAM-1. *Nouv. Rev. Fr. Hematol.* **34** (suppl), S9–13 (1992).
20. Sato, T. N., Qin, Y., Kozak, C. A. & Audus, K. L. Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system. *Proc. Natl Acad. Sci. USA* **90**, 9355–9358 (1993).
21. Weibel, E. R. & Palade, G. E. New cytoplasmic components in arterial endothelia. *J. Cell Biol.* **23**, 101–112 (1964).
22. Grant, D. S., Lelkes, P. I., Fukuda, K. & Kleinman, H. K. Intracellular mechanisms involved in basement membrane induced blood vessel differentiation *in vitro*. *In Vitro Cell. Dev. Biol.* **27A**, 327–336 (1991).
23. Bjornson, C. R., Rietze, R. L., Reynolds, B. A., Magli, M. C. & Vescovi, A. L. Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells *in vivo*. *Science* **283**, 534–537 (1999).
24. Clarke, D. L. *et al.* Generalized potential of adult neural stem cells. *Science* **288**, 1660–1663 (2000).
25. LaBarge, M. A. & Blau, H. M. Biological progression from adult bone marrow to mononucleate muscle stem cell to multinucleate muscle fiber in response to injury. *Cell* **111**, 589–601 (2002).
26. Chen, T. R. Cytogenetics of somatic cell hybrids. I. Progression of stemlines in continuous uncloned cultures of man–mouse cell hybrids. *Cytogenet. Cell Genet.* **23**, 221–230 (1979).
27. Schall, D. & Rechsteiner, M. Kinetics of human chromosome loss from 3T3–human hybrid cells. *Somatic Cell Genet.* **4**, 661–676 (1978).
28. Vacquier, V. D. The fertilizing capacity of sea urchin sperm rapidly decreases after induction of the acrosome reaction. *Dev. Growth Differ.* **21**, 61–69 (1979).
29. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
30. Palmer, T. D., Markakis, E. A., Willhoite, A. R., Safar, F. & Gage, F. H. Fibroblast growth factor-2 activates a latent neurogenic program in neural stem cells from diverse regions of the adult CNS. *J. Neurosci.* **19**, 8487–8497 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. G. Farquhar and R. V. Stan for critically evaluating Weibel–Palade bodies; D. Chambers for conducting FACS analysis; J. Simon for help in constructing figures; L. Moore, B. Miller, A. Hyunh and S. Thuret for technical assistance; J. D. Esko's laboratory for advice; V. A. Sciorra and M. L. Gage for critical reading of the manuscript. This work was supported by the National Institutes of Health (National Institute on Aging) and the Lookout Fund. A.E.W. is a Damon Runyon fellow supported by The Damon Runyon Cancer Research Foundation. K.N. was supported by a JSPS Postdoctoral Fellowship for Research Abroad. F.H.G. is supported by the Vi and John Adler Chair.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to F.H.G. (gage@salk.edu).

The homeotic protein AGAMOUS controls microsporogenesis by regulation of *SPOROCTELESS*

Toshiro Ito¹, Frank Wellmer¹, Hao Yu^{1,2}, Pradeep Das¹, Natsuko Ito¹, Márcio Alves-Ferreira^{1,3}, José Luis Riechmann¹ & Elliot M. Meyerowitz¹

¹Division of Biology 156-29, California Institute of Technology, 1200 E. California Blvd, Pasadena, California 91125, USA

²Department of Biological Sciences, Faculty of Science, National University of Singapore, 10 Science Drive 4, 117543, Singapore

³Department of Genetics, Federal University of Rio de Janeiro, Av. Pau Brasil 211 A2-76, 21941-590, Rio de Janeiro, Brazil

The *Arabidopsis* homeotic gene *AGAMOUS* (*AG*) is necessary for the specification of reproductive organs (stamens and carpels) during the early steps of flower development^{1–3}. *AG* encodes a transcription factor of the MADS-box family that is expressed in stamen and carpel primordia. At later stages of development, *AG* is expressed in distinct regions of the reproductive organs^{2–5}. This suggests that *AG* might function during the maturation of stamens and carpels, as well as in their early development. However, the developmental processes that *AG* might control during organogenesis and the genes that are regulated by this factor are largely unknown. Here we show that microsporogenesis, the process leading to pollen formation, is induced by *AG* through activation of the *SPOROCTELESS* gene (*SPL*, also known as *NOZZLE/NZZ*), a regulator of sporogenesis^{6,7}. Furthermore, we demonstrate that *SPL* can induce microsporogenesis in the absence of *AG* function, suggesting that *AG* controls a specific process during organogenesis by activating another regulator that performs a subset of its functions.

AG triggers reproductive organ development^{8,9}, a process which presumably involves temporally and spatially specific expression of a large number of genes¹⁰. However, only a few putative target genes of *AG* are currently known^{11–13}. One of these genes is the MADS-box gene *SHATTERPROOF2* which shares partially redundant functions with *AG* in addition to its function in fruit dehiscence^{12,14,15}. No other putative target genes have been identified that may function downstream of *AG* during floral organ formation.

To analyse the processes downstream of *AG*, we produced a strain with inducible *AG* activity. This strain is homozygous for the *ag-1* null mutation and transgenic for 35S::*AG-GR*, a constitutive promoter-driven *AG* gene with a carboxy-terminal fusion to the steroid-binding domain of the rat glucocorticoid receptor^{16,17}. Plants of this genotype make *ag-1* mutant flowers, with indefinite numbers of successive whorls of sepal–petal–petal (Fig. 1a). Continuous treatment with the steroid hormone dexamethasone (DEX)—which leads to a translocation of the fusion protein from

the cytoplasm to the nucleus—resulted in flowers with functional stamens and carpels, which resemble the flowers of 35S::*AG* plants⁸ (Fig. 1b). Thus, the *AG-GR* fusion protein can perform the function of *AG*. In contrast, a single DEX treatment resulted in only a partial phenotypic rescue; flowers bore petals with stamenoid structures (Fig. 1c). Most of the petals produced in whorl 3 of these flowers had locules (pollen sacs) at the lateral edges containing normal-looking pollen grains (Fig. 1d–f). These results show that a single DEX treatment is not enough to produce normal stamens, but is sufficient to induce microsporogenesis.

The strain with inducible *AG* activity was used in a gene expression profiling experiment performed with a complementary DNA microarray¹⁰. Several *AG*-responsive genes were identified, one of which was *SPL* (see Supplementary Fig. S1). *SPL* encodes a putative transcription factor, and has been implicated in ovule pattern formation and early microsporogenesis^{6,7,18}. In *spl* mutants, the differentiation of primary sporogenous cells as well as anther wall formation are blocked. This results in mutant anthers that lack pollen grains^{6,7}. The induction of *SPL* by *AG* activity was confirmed by semi-quantitative polymerase chain reaction with reverse transcription (RT–PCR), as well as real-time PCR (Fig. 2a; see also Supplementary Fig. S2). *AG*-dependent *SPL* upregulation was detected at the 4 hour time-point, and the signal became stronger at 8 h, remaining relatively stable afterwards. We next tested whether *AG* activates *SPL* in the presence of 10 μ M cycloheximide, an efficient inhibitor of protein synthesis^{17,19}. Cycloheximide did not block the induction of *SPL* by *AG*, suggesting that this induction might be direct (Fig. 2a).

To examine whether there are putative binding sites of *AG* in the *SPL* promoter, the *SPL* genomic region was examined for the 16-base pair consensus binding sequence of *AG*^{20,21}, which contains the 10-bp ‘CARG-box core’ (5′-CCNN(A/T)₄GG-3′). In the 3′ region of the gene (897 bp downstream from the stop codon), one site was identified that contains a single mismatch from the consensus sequence of the CARG box (Fig. 2b). An electrophoretic

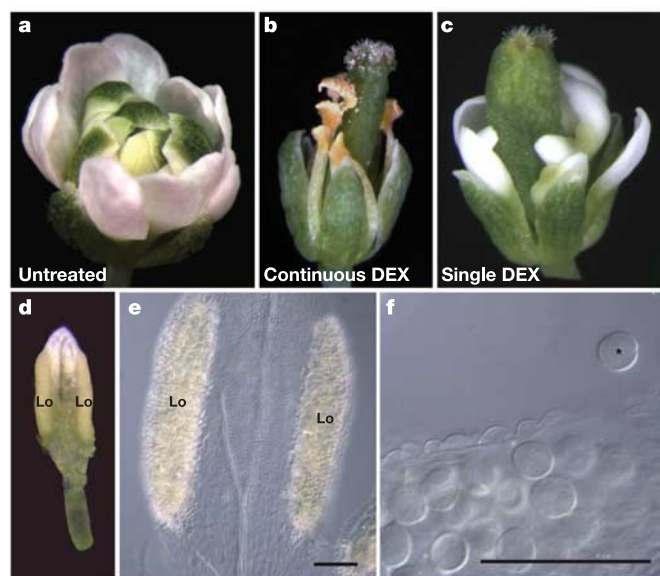


Figure 1 Activation of *AG-GR* rescues the *ag-1* mutant phenotype. **a–c**, Flowers from *ag-1* 35S::*AG-GR* plants untreated (**a**), or after continuous (**b**), or a single (**c**) DEX treatment. Treatments were initiated 12 days before imaging. Four DEX treatments at 0.5-day intervals, or three treatments at 2-day intervals, were sufficient to fully rescue the mutant phenotypes (data not shown). **d**, Petal with loculed structures produced in whorl 3 of the flower shown in **c**. **e, f**, Close-up of a loculed petal after clearing. Morphologically normal pollen grains (**f**) were produced in the locules (**e**). Asterisk indicates a pollen grain released by manually breaking a locule. Lo, locules. Scale bars, 100 μ m.

mobility shift assay using a synthetic DNA probe showed that recombinant AG protein binds to this sequence *in vitro* (Fig. 2c). The binding was increased when the single mismatch of the wild-type sequence was altered to produce the consensus sequence, or lost when additional mutations were introduced into the CARG-box core (Fig. 2b, c). Another site with a perfect CARG-box core, but with three mismatches to the 16-bp AG consensus sequence, was found 100 bp downstream from the stop codon. No clear binding of AG to this site was observed *in vitro* (data not shown).

To evaluate whether the AG binding site detected in the 3' region is necessary for the AG-dependent induction of *SPL*, translational fusions between *SPL* and the *GUS* reporter gene were made and transformed into wild-type plants. Overall, *GUS* expression patterns in plant transgenic lines containing the wild-type *SPL*-*GUS* reporter gene recapitulated the endogenous *SPL* messenger RNA expression pattern⁶ (Fig. 2d-h, see also Supplementary Fig. S3a-h). *GUS* expression started to appear at the lateral edges of the stamen primordia in stage 6 floral buds (Fig. 2e), and subsequently expanded, covering the developing stamens during

stages 6-8 (Fig. 2f-h). In addition to the continued *GUS* expression in the developing stamens, staining was also observed in ovule primordia at stage 10, as well as in the nucellus, the integuments and the funiculus of developing ovules at stages 11-13 (see Supplementary Fig. S3a-g). In contrast, transgenic plants carrying a *SPL*-*GUS* fusion in which the putative AG-binding site had been mutated (leading to a loss of AG binding *in vitro*; see above), showed weaker overall *GUS* activity and reduced expression domains (Fig. 2i-l; see also Supplementary Fig. S3i-k). During stages 6-8, weak staining was observed at only the lateral sides of the stamen primordia, and not throughout the primordia as observed using the wild-type construct (Fig. 2i-l). In the ovules of the transgenic lines containing the mutated construct, staining was observed in the nucellus, and very weak staining was observed in the inner integuments; however, staining was not observed in the outer integuments or in the funiculus at stages 11-13 (compare Supplementary Fig. S3i-k with e-g). This loss of *GUS* staining in the integuments and the funiculus corresponds to the pattern of AG expression in developing ovules; AG is specifically expressed in the integuments and

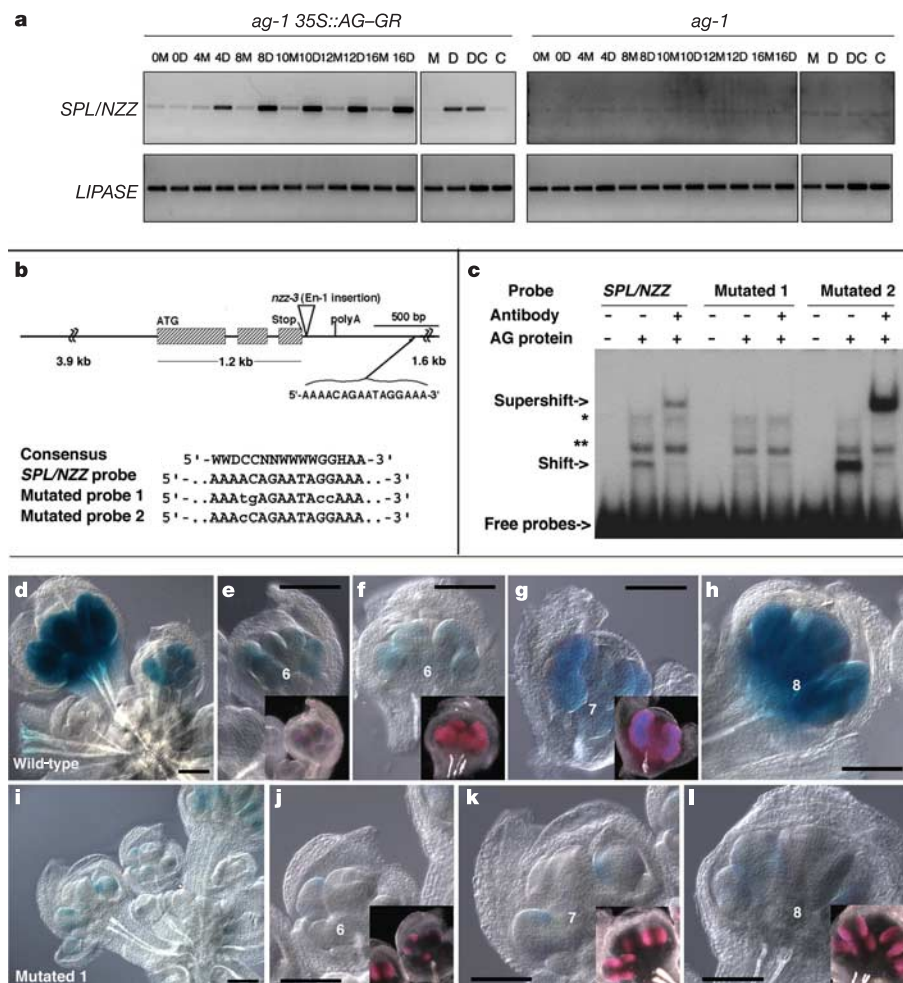


Figure 2 AG-mediated induction of *SPL* through a binding site in the 3' region of the gene. **a**, RT-PCR for *SPL* (top) or for a ubiquitously expressed lipase gene (bottom) using RNA samples isolated from *ag-1* 35S::AG-GR plants (left) or control *ag-1* plants (right) 0, 4, 8, 10, 12 or 16 h after mock (M) or DEX (D) treatment. The smaller panels show the results for samples 4 h after simultaneous treatment with DEX and 10 μM cycloheximide (DC) or cycloheximide alone (C). **b**, Diagram of the *SPL* genomic region (top), and the 16-bp consensus binding sequence of AG (bottom). The CARG-box-like sequence found 897 bp downstream from the stop codon of the *SPL* gene, as well as the mutations that were introduced in it (indicated with small letters) are shown. D, not C; H, not G; N, A/T/G/C; W, A/T. **c**, Electrophoretic mobility shift assay using recombinant AG protein

tagged with the c-Myc epitope, anti-c-Myc monoclonal antibody (9E10) and the 30-bp oligonucleotide probes described in **b**. Single and double asterisks indicate nonspecific binding from the bacterial extract. **d-l**, *GUS* expression patterns in transgenic lines carrying *SPL*-*GUS* fusion constructs either with the wild-type (**d-h**) or with a mutated (**i-l**) CARG-box-like sequence. The mutation was the same as for the mutated probe 1 shown in **b**. *GUS* staining for both constructs was performed under the same conditions. At least ten independent lines for each construct were analysed and representative images are shown. Insets in **e-g** and **j-l** are dark-field images showing enhanced staining of half-magnifications. Numbers indicate floral stages³⁰. Scale bars, 100 μm.

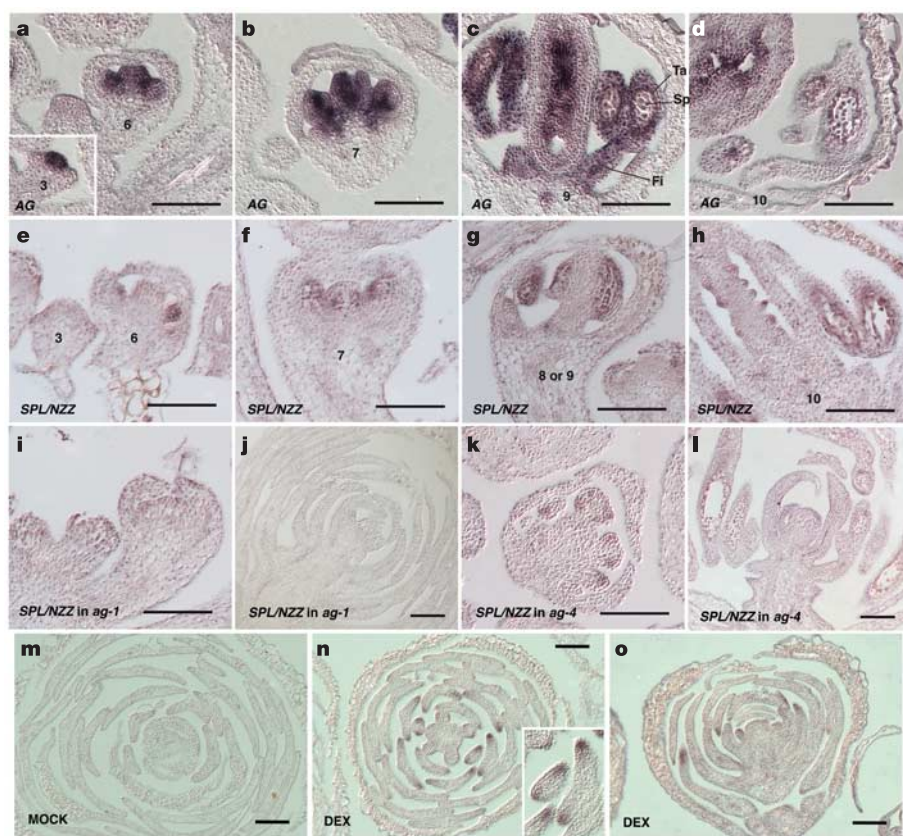


Figure 3 Expression of *SPL* and *AG* in wild-type flowers, and expression of *SPL* in flowers of *ag* mutants and in the lines with inducible *AG* activity. **a–h**, Expression of *AG* (**a–d**) and *SPL* (**e–h**) in wild-type flowers. **i–l**, Expression of *SPL* in *ag-1* (**i, j**) and *ag-4* (**k, l**) mutant flowers. **m–o**, *SPL* is induced by *AG* activity at the lateral edges of the distal part of organ primordia. Inflorescences of *ag-1 35S::AG-GR* plants were fixed 8 h after mock- (**m**) or

DEX (**n, o**) treatment. Transverse (**m, n**) and longitudinal (**o**) sections were hybridized with a *SPL*-specific antisense probe. The inset in **n** shows the close-up view of the *SPL* signals. Numbers indicate floral stages³⁰. Fi, filament; Sp, sporogenous cell; Ta, tapetum. Scale bars, 100 μ m.

the funiculus of stage 11–12 flowers⁵. These results show that the CARG-box-like sequence in the 3' region of the *SPL* gene, which is bound by *AG* *in vitro*, is necessary for normal *SPL* expression in developing stamens and ovules. The importance of the 3' region of *SPL* for its function is further supported by the null *nzz-3* allele, which is caused by the insertion of a transposon 20-bp downstream from the stop codon⁶ (Fig. 2b). Together, these data suggest that *AG* may bind to the 3' region of the *SPL* gene and thus may directly regulate *SPL* expression during the development of the reproductive organs.

To determine whether *AG* is responsible for *SPL* expression throughout stamen development, the expression patterns of *SPL* and *AG* were compared in detail. The expression of *SPL* in stamen primordia during stages 6–7 was completely within the expression domain of *AG* (compare Fig. 3a, b with e, f). At stages 8–9, the *SPL* transcript was localized to tapetal and sporogenous cells (microspore mother cells)⁶, whereas there was strong *AG* expression in the tapetum and the filament, but not in the sporogenous cells³ (Fig. 3c, g). In anthers of stage 10 flowers, *SPL* expression continued in the tapetum and microspores⁶, whereas *AG* was expressed in the tapetum, but not in microspores³ (Fig. 3d, h). These partly overlapping expression patterns are in agreement with the idea that *AG* is required to activate *SPL* expression, but is not necessary for its maintenance. In the weak *ag-4* mutant, which produces flowers with stamens in whorl 3 (ref. 22), weak but distinct expression of *SPL* was observed in stamen primordia, but expression was barely detectable in sporogenous cells and inner organs of mature flowers (Fig. 3k, l). No clear expression of *SPL* was observed in the strong

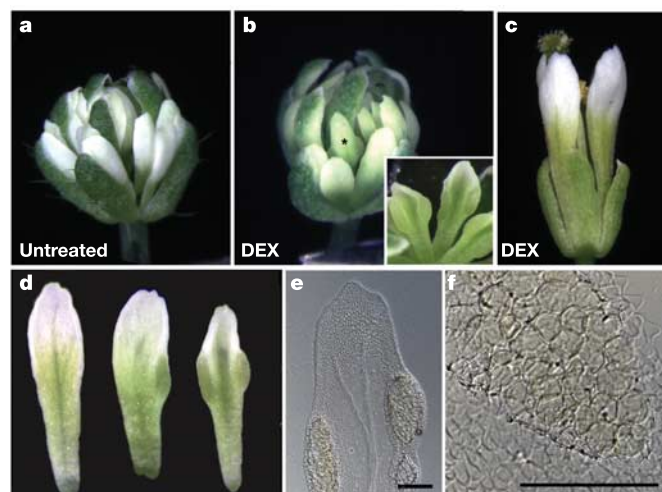


Figure 4 *SPL* is sufficient to induce microsporogenesis in the inner organs of *ag-1* mutant flowers. **a, b**, Flowers from *ag-1 35S::SPL-GR* transgenic plants untreated (**a**) or after continuous DEX treatment (**b**). The inset in **b** shows loculed petals produced in the inner whorls. DEX-treated non-transgenic *ag-1* plants did not show any loculed petals (data not shown). **c**, Flower from a *35S::SPL-GR* plant (wild-type background) after continuous DEX treatment. **d**, Normal-looking petals in whorl 2 (left) and loculed petals (middle and right) that developed in the inner whorls of a DEX-treated *ag-1 35S::SPL-GR* flower. **e, f**, Cleared view of loculed petals induced by ectopic *SPL* activity. Asterisk in **b** indicates an internal loculed petal. Scale bar, 100 μ m.

ag-1 mutant allele (Fig. 3i, j), indicating that the induction of *SPL* expression is dependent on AG.

To obtain further evidence for the induction of *SPL* by AG, localization of *SPL* transcripts in flowers of *ag-1 35S::AG-GR* plants was examined. Eight hours after DEX treatment, induction of *SPL* was observed along the lateral edges of the distal parts of organ primordia, which matches the region where microsporogenesis is induced in stamoid petals (compare Fig. 3m–o with Fig. 1d). This expression domain also coincides with the initial expression of *SPL* observed in the *SPL-GUS* reporter lines (Fig. 2e). No induction was observed in fully developed sepals or petals (Fig. 3n, o). These results demonstrate that AG specifically upregulates *SPL* expression in a developmentally controlled manner in specific floral tissues.

To test whether *SPL* can induce microsporogenesis in the absence of AG function, a *35S::SPL-GR* construct was used. The functionality of the *SPL-GR* fusion protein during microsporogenesis was confirmed in the *nzz-2* mutant background (data not shown). In transgenic lines in the *ag-1* background, loculed petals were observed after continuous DEX treatment (Fig. 4a, b). These loculed structures were similar to the ones induced by a single DEX treatment in *ag-1 35S::AG-GR* plants (compare Fig. 4b, e with Fig. 1d, e), and inside the locules, pollen grains were observed (Fig. 4f). This result shows that *SPL* can induce microsporogenesis in the absence of AG activity, suggesting that *SPL* acts as a central mediator for the function of AG in sporogenesis. Although *SPL-GR* was ectopically expressed, locule formation was observed in *ag-1* mutant flowers only at the lateral edges of organs in whorl 3 and inward; the ectopic locule formation was never observed in whorl 2 of DEX-treated flowers (Fig. 4c, d). These data suggest that other whorl-specific factors are involved in the induction of microsporogenesis.

Our results, together with the previous identification of *NAP* (an NAC-family transcription factor gene) as a late-stage target of *APETALA3* (ref. 17), indicate that floral homeotic proteins control target genes with various functions throughout organ development. In addition, the interaction of AG and *SPL* shows the likelihood that AG activates other regulators that perform a subset of its functions. Therefore, this study provides direct evidence to support the idea that plant homeotic genes control organogenesis as part of a regulatory hierarchy. □

Methods

Plant materials and treatments

All plants used in this study were in the Landsberg *erecta* background, and were grown at 22 °C under constant illumination. Transgenic plants were generated by *Agrobacterium*-mediated infiltration. The *35S::AG-GR* construct was transformed into wild-type plants and the primary T1 transformants were screened by kanamycin selection. A line carrying a single transgene insertion, which showed the AG-overexpression phenotype⁶ in a DEX-dependent fashion, was subsequently crossed into the *ag-1* background. The *35S::SPL-GR* construct was transformed directly into *ag-1/+* plants, and the T1 plants were screened by Basta selection. Phenotypic analyses in the *ag-1* background were performed in the T1 generation, as well as in the T2 generation that was produced by selfing *ag-1/+* T1 plants. One transgenic line for *35S::SPL-GR* was crossed to *nzz-2/+* plants to obtain lines in the *nzz-2* mutant background in subsequent generations. The *SPL-GUS* constructs were transformed into wild-type plants. GUS staining was performed on the T1 plants screened by Basta selection. For continuous DEX treatment of the *ag-1 35S::AG-GR* or *ag-1 35S::SPL-GR* plants, the plants were watered with 10 μM DEX daily. For a single DEX treatment, the inflorescences were dipped in water containing 10 μM DEX and 0.015% Silwet L-77 (ref. 19). Cleared whole-mount observation was performed according to published protocols²³.

Plasmid constructs

The *35S::AG-GR* construct was produced as follows: the coding region of AG was amplified from inflorescence cDNA using primers P1 and P2 (see Supplementary Table 1) and cloned into a blunt-ended *EcoRI* site of pBluescript SK (Stratagene) to produce pSK-AG. To produce pSK-AG-GR, pSK-AG was digested with *EcoRI*, filled-in and ligated with a DNA fragment containing a rat glucocorticoid hormone binding domain¹⁶, which was excised with *Bam*HI and *Xba*I from pRS020 (pDeltaGRBX)¹⁷ and filled in. The AG-GR fragment was released from pSK-AG-GR by digestion with *Xba*I and *Clal*, and ligated into the corresponding sites of the binary vector pMAT137 (a derivative of pMAT037; ref. 24) that contains tandem cauliflower mosaic virus 35S enhancers and a terminator. For the reporter gene constructs, a *SPL* 6,646-bp genomic fragment (positions 4,303–10,958 of

BAC clone F27G19 (GenBank accession number AL078467)) comprising *SPL* and the flanking intergenic sequences was amplified using UltraPfu High-Fidelity DNA polymerase (Stratagene) with primers P3 and P4 (see Supplementary Table 1) and wild-type Columbia genomic DNA. The product was cloned into the pCR II vector (Invitrogen) to produce pCR-SPLG. pCR-SPLG was mutagenized in the 3' CARG-box-like sequence (897-bp downstream from the stop codon) by using the Quikchange II XL site-directed mutagenesis kit (Stratagene) with primers P5 and P6 (see Supplementary Table 1) to produce pCR-SPLG-M3. Both pCR-SPLG and pCR-SPLG-M3 were digested with *Bsp*EI (6939 of F27G19), filled in and then ligated with a *Sfo*I-digested *GUS* fragment²⁵ to produce pCR-SPLG-ori-GUS and pCR-SPLG-M3-GUS, respectively. The *SPL-GUS* reporter fragments from pCR-SPLG-ori-GUS and pCR-SPLG-M3-GUS were cut out with *Asp*718 (later filled in) and *Not*I, and cloned into the binary vector pMIBart (a derivative of pART²⁶) digested with *Ecl*136 and *Not*I. The *35S::SPL-GR* construct was produced as follows: *SPL* cDNA was amplified using primers P7 and P8 (see Supplementary Table 1), and cloned into the pCRII vector (Invitrogen) to produce pCR-SPL. To produce *35S::SPL-GR*, pCR-SPL was digested with *Mlu*I and *Xho*I, and the resulting insert was cloned into the pGreen0236TI vector, which is a derivative of pGreen0229 (ref. 27) containing a 35S promoter with tandem enhancers and the hormone binding domain of the rat glucocorticoid receptor followed by a 6 × haemagglutinin tag. The construction of the pG0236TI vector will be described elsewhere (T.I. and E.M.M., manuscript in preparation).

Microarray, RT-PCR and real-time PCR analysis

In the time course experiments after induction of AG activity, inflorescences containing floral buds of stages 1–10 were collected from *ag-1 35S::AG-GR* plants, the stems of which were about 5-cm long, 0, 4, 8, 10, 12 and 16 h after a single mock- or DEX treatment. Total RNA was isolated from tissue samples using Tri-reagent (Molecular Research Center Inc.) or the RNeasy Mini Kit (Qiagen). Two sets of independent RNA samples from *ag-1 35S::AG-GR* plants and two sets of RNA samples from the *ag-1* single mutant were prepared. Microarray experiments were performed as previously described^{10,28}. For RT-PCR, total RNA was reverse-transcribed using the ThermoScript RT-PCR system (Invitrogen). As semi-quantitative RT-PCR, the expression of *SPL* (At4g27330) and a control lipase gene (At1g10740) were analysed using primers P9 and P10 after 30 cycles and primers P11 and P12 after 24 cycles, respectively (see Supplementary Table 1). The gel was blotted onto Hybond-N membrane and hybridized with ³²P-labelled *SPL* or lipase RT-PCR products. Real-time PCR was performed with the SYBR Green PCR Master Mix (Applied Biosystems) using the Applied Biosystems GeneAmp 5700 sequence detection system according to the manufacturer's instructions. The sequences of *SPL* primers P13 and P14 and actin control primers P15 and P16 used for real-time PCR are shown in Supplementary Table 1.

In vitro DNA binding assay

Electrophoretic mobility shift assays (EMSA) were performed as described²⁰. Sequences of oligonucleotides for probes are shown in Supplementary Table 1. For the CARG-box-like sequence located 897-bp downstream from the stop codon (*SPL* probe; Fig. 2b, c), oligonucleotides O17 and O18 were used. Mutated versions of this probe were prepared with oligonucleotides O19 and O20 (mutated probe 1), and oligonucleotides O21 and O22 (mutated probe 2). For the site with a perfect CARG-box core located 100-bp downstream from the stop codon, oligonucleotides O23 and O24 were used. For supershift assays, a monoclonal anti-c-Myc antibody (9E10, Santa Cruz Biotechnology) was used²⁰.

In situ hybridization and GUS staining

Non-radioactive *in situ* hybridizations were performed as previously described²⁹. To produce a *SPL*-specific anti-sense probe, the RT-PCR product was cloned into the pCRII vector (Invitrogen) and used as a template for *in vitro* transcription. The AG-specific probe was synthesized from plasmid pCIT565(AG)⁴. GUS staining was performed as previously described²⁵.

Received 9 May; accepted 7 June 2004; doi:10.1038/nature02733.

1. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52 (1989).
2. Yanofsky, M. F. *et al.* The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* **346**, 35–39 (1990).
3. Bowman, J. L., Drews, G. N. & Meyerowitz, E. M. Expression of the *Arabidopsis* floral homeotic gene *AGAMOUS* is restricted to specific cell types late in flower development. *Plant Cell* **3**, 749–758 (1991).
4. Drews, G. N., Bowman, J. L. & Meyerowitz, E. M. Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991–1002 (1991).
5. Reiser, L. *et al.* The *BELL1* gene encodes a homeodomain protein involved in pattern formation in the *Arabidopsis* ovule primordium. *Cell* **83**, 735–742 (1995).
6. Schiefthaler, U. *et al.* Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 11664–11669 (1999).
7. Yang, W. C., Ye, D., Xu, J. & Sundaresan, V. The *SPOROCTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. *Genes Dev.* **13**, 2108–2117 (1999).
8. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71**, 119–131 (1992).
9. Honma, T. & Goto, K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* **409**, 525–529 (2001).
10. Wellmer, F., Riechmann, J. L., Alves-Ferreira, M. & Meyerowitz, E. M. Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell* **16**, 1314–1326 (2004).
11. Gustafson-Brown, C., Savidge, B. & Yanofsky, M. F. Regulation of the *Arabidopsis* floral homeotic gene *APETALA1*. *Cell* **76**, 131–143 (1994).

12. Savidge, B., Rounsley, S. D. & Yanofsky, M. F. Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell* **7**, 721–733 (1995).
13. Lenhard, M., Bohnert, A., Jurgens, G. & Laux, T. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between *WUSCHEL* and *AGAMOUS*. *Cell* **105**, 805–814 (2001).
14. Liljgren, S. J. *et al.* SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766–770 (2000).
15. Pinyopich, A. *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85–88 (2003).
16. Lloyd, A. M., Schena, M., Walbot, V. & Davis, R. W. Epidermal cell fate determination in *Arabidopsis*: patterns defined by a steroid-inducible regulator. *Science* **266**, 436–439 (1994).
17. Sablowski, R. W. & Meyerowitz, E. M. A homolog of *NO APICAL MERISTEM* is an immediate target of the floral homeotic genes *APETALA3/PISTILLATA*. *Cell* **92**, 93–103 (1998).
18. Balasubramanian, S. & Schneitz, K. *NOZZLE* links proximal-distal and adaxial-abaxial pattern formation during ovule development in *Arabidopsis thaliana*. *Development* **129**, 4291–4300 (2002).
19. Wagner, D., Sablowski, R. W. & Meyerowitz, E. M. Transcriptional activation of *APETALA1* by *LEAFY*. *Science* **285**, 582–584 (1999).
20. Shiraishi, H., Okada, K. & Shimura, Y. Nucleotide sequences recognized by the *AGAMOUS* MADS domain of *Arabidopsis thaliana* in vitro. *Plant J.* **4**, 385–398 (1993).
21. Huang, H., Mizukami, Y., Hu, Y. & Ma, H. Isolation and characterization of the binding sequences for the product of the *Arabidopsis* floral homeotic gene *AGAMOUS*. *Nucleic Acids Res.* **21**, 4769–4776 (1993).
22. Sieburth, L. E., Running, M. P. & Meyerowitz, E. M. Genetic separation of third and fourth whorl functions of *AGAMOUS*. *Plant Cell* **7**, 1249–1258 (1995).
23. Schneitz, K., Hülskamp, M. & Pruitt, R. E. Wild-type ovule development in *Arabidopsis thaliana*: a light microscope study of cleared whole-mount tissue. *Plant J.* **7**, 731–749 (1995).
24. Matsuoka, K. & Nakamura, K. Propeptide of a precursor to a plant vacuolar protein required for vacuolar targeting. *Proc. Natl Acad. Sci. USA* **88**, 834–838 (1991).
25. Ito, T., Sakai, H. & Meyerowitz, E. M. Whorl-specific expression of the *SUPERMAN* gene of *Arabidopsis* is mediated by *cis* elements in the transcribed region. *Curr. Biol.* **13**, 1524–1530 (2003).
26. Gleave, A. P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol. Biol.* **20**, 1203–1207 (1992).
27. Hellens, R. P., Edwards, E. A., Leyland, N. R., Bean, S. & Mullineaux, P. M. pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol. Biol.* **42**, 819–832 (2000).
28. Wagner, D. *et al.* Floral induction in tissue culture: a system for the analysis of *LEAFY*-dependent gene regulation. *Plant J.* **39**, 273–282 (2004).
29. Long, J. A. & Barton, M. K. The development of apical embryonic pattern in *Arabidopsis*. *Development* **125**, 3027–3035 (1998).
30. Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755–767 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Chow for experimental help; M. Hasebe; former Meyerowitz laboratory members, J. A. Long, K. Goto, M. Frohlich and P. Kumar; and current laboratory members, especially P. Sieber, I. Negruțiu, A. Dubois, C. Ohno, C. Baker, E. Haswell, M. Heisler, V. R. Gonehal and Y. X. Zhao for comments on the manuscript. M.A.-F. is indebted to IPBO and M. Van Montagu for a fellowship sponsored by Aventis CropSciences. This work was supported by a grant from the National Institutes of Health to E.M.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.M. (meyerow@its.caltech.edu).

First cleavage plane of the mouse egg is not predetermined but defined by the topology of the two apposing pronuclei

Takashi Hiragi & Davor Solter

Max-Planck Institute of Immunobiology, D-79108 Freiburg, Germany

Studies of experimentally manipulated embryos^{1–4} have led to the long-held conclusion that the polarity of the mouse embryo remains undetermined until the blastocyst stage. However, recent studies^{5–7} reporting that the embryonic–abembryonic axis of the blastocyst arises perpendicular to the first cleavage

plane, and hence to the animal–vegetal axis of the zygote, have led to the claim that the axis of the mouse embryo is already specified in the egg. Here we show that there is no specification of the axis in the egg. Time-lapse recordings show that the second polar body does not mark a stationary animal pole, but instead, in half of the embryos, moves towards a first cleavage plane. The first cleavage plane coincides with the plane defined by the two apposing pronuclei once they have moved to the centre of the egg. Pronuclear transfer experiments confirm that the first cleavage plane is not determined in early interphase but rather is specified by the newly formed topology of the two pronuclei. The microtubule networks that allow mixing of parental chromosomes before dividing into two may be involved in these processes.

Polarity formation of the mammalian preimplantation embryo remains a controversial subject. Because of their highly regulative capacity in coping with experimental manipulations (for example, blastomere isolation^{1–3} and chimaera formation⁴) mouse embryos have long been thought to lack polarity until the blastocyst stage. Recently this assumption has been challenged in two independent approaches^{6,7} showing that the embryonic–abembryonic (Em–Ab) axis of the mouse blastocyst is perpendicular to the first cleavage plane. Considering the second polar body (2pb) as a stationary marker of the animal pole during preimplantation development, the authors concluded that polarity of the mouse embryo is specified in the egg, as it is for most non-mammalian animals. Possible movement of the 2pb was dismissed as ‘occasional’^{5,6}. It has also been proposed that the plane of initial cleavage passes through both the sperm entry position (SEP) and the 2pb, and that the SEP can define the future Em–Ab axis of the embryo^{8,9}. However, our preliminary

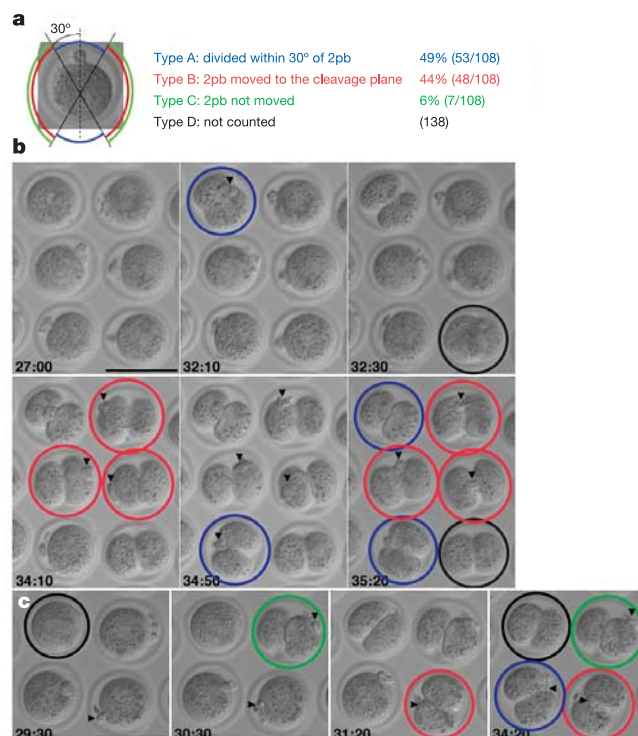


Figure 1 The 2pb moves towards the first cleavage plane in half the embryos. **a**, Embryos are classified into four types: type A (blue), divided within 30° of the 2pb; type B (red), the 2pb moved towards the first cleavage plane; type C (green), the 2pb did not reach the cleavage plane; type D (black), not counted due to the unclear topological relationship. **b, c**, Sequential DIC images of embryos cultured *in vitro*. The embryos assigned to each type are marked with coloured circles. Arrowheads indicate the 2pb. In each frame, time is given in h:min after hCG injection. Scale bars represent 100 μm.

analysis demonstrated very dynamic movement of the 2pb towards the first cleavage plane before and after the first division in a substantial number of embryos, contrasting directly with the notion that the first cleavage plane is meridional, defined by the position of a stationary 2pb and coincidental with the putative animal–vegetal (A–V) axis of the egg.

To investigate this discrepancy further, we set up a time-lapse recording system for the mouse preimplantation embryo (see Methods). After time-lapse recording of embryos (for 24 h, starting from the zygote) and a further two days of *in vitro* culture, 92% (185 out of 202) of the embryos had developed to the blastocyst stage, indicating that our observation conditions closely reflect the situation *in vivo*. We then classified the recorded embryos into four categories (Fig. 1a): type A, in which cleavage occurs within 30° of the 2pb; type B, in which the 2pb moves towards the cleavage plane; type C, in which the 2pb does not reach the cleavage plane; and type D, in which the extent of 2pb movement could not be reliably judged because either the 2pb was not clearly visible or the cleavage plane was close to parallel to the observation plane thus making determination of their relationship difficult or impossible. The type D group, representing more than half of the embryos, was excluded from the analysis. Figure 1b and c shows examples of each category (full sequence in Supplementary Movies 1 and 2, respectively). In a substantial number of embryos (circled in red in Fig. 1b), the 2pb was separate from the first cleavage plane at division (see $t = 34:10$), gradually moving towards the cleavage plane thereafter (at $t = 34:50$ and $35:20$). In some embryos (circled in red in Fig. 1c), the 2pb began moving even before cleavage. In total, only 49% of the embryos (53 out of 108, type A) were consistent with the assertion that the cleavage plane is essentially aligned with the A–V axis of the zygote, whereas 51% (55 out of 108, type B plus C) were not. In light of such a substantial population forming the first cleavage plane without any relationship to the 2pb position, we conclude that if there is a predetermined A–V axis in the mouse egg, 2pb does not mark its position nor is there any discernible relationship between it and the plane of the first division. Therefore without a stable

morphological feature defining the A–V axis, the entire concept of a predetermined A–V axis in the mammalian egg should be discarded.

To address whether formation of the first cleavage plane is a precisely defined event or, instead, completely random, we performed a more detailed analysis using *in vitro*-fertilized embryos collected just after the formation of a fertilization cone. Whereas the first cleavage plane passed close to the 2pb of the zygote in some embryos (type A; Fig. 2a), many showed the cleavage plane separate from the 2pb (type B and C; Fig. 2b, c). Interestingly, in all three types, the cleavage plane was always the same as the plane separating the two apposing pronuclei in the centre of the egg (Fig. 2d), that is, the cleavage plane was perpendicular to the line going through the two apposing pronuclei (Fig. 2a–c; and the corresponding Supplementary Movies 3–5). Once the cleavage plane was specified, the 2pb moved towards the cleavage furrow before and after cleavage (Fig. 2b; and Supplementary Movie 4). In rare cases, the 2pb remained separate from the cleavage plane at the two-cell stage (type C; Fig. 2c). Out of 25 embryos observed in this series, the cleavage plane in 18 embryos (72%) could be predicted at fertilization, as the paths of pronuclear formation and movement were usually the shortest way to the centre for both the male and female pronuclei (both male and female pronuclei form in the egg periphery and move in a microtubule-dependent manner towards the centre^{10,11}). In those embryos the cleavage plane defined by the two apposing pronuclei also bisects the shortest arc between the 2pb and the SEP (fertilization cone). In three embryos (12%), the topological relationship of the pronuclei changed by gradual rotation in an apparently random way during and after pronuclei movement to the centre, and the cleavage plane eventually formed according to the final apposing plane of the two pronuclei just before pronuclear membrane breakdown (PNMBD) (data not shown). In the remaining four embryos (16%) the cleavage plane was so far from perpendicular to the observation plane that its relationship with the plane of nuclear apposition could not be determined. These data suggest that the first cleavage plane is

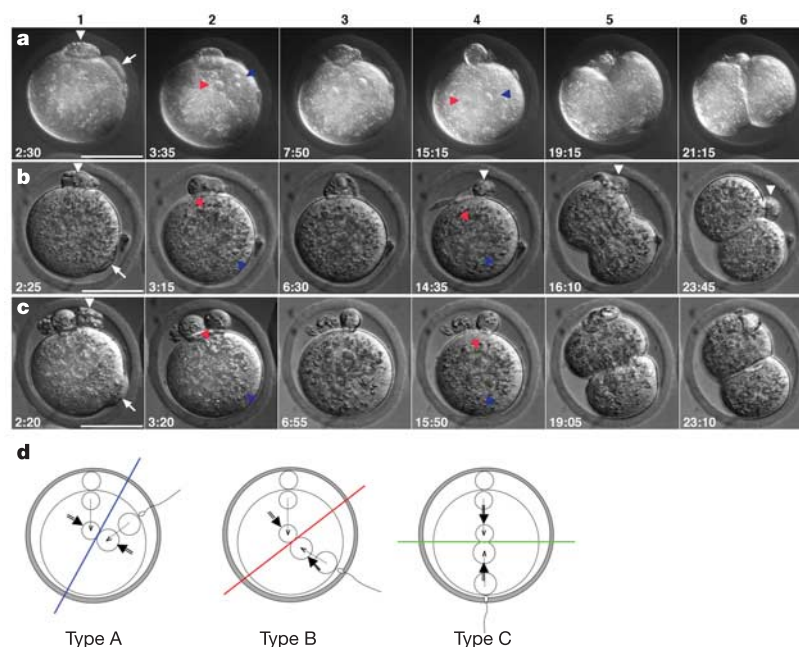


Figure 2 The first cleavage plane is specified by the topology of the two apposing pronuclei in the egg centre. **a–c**, Sequence for an embryo of each type. (1) Fertilization cone (white arrow) formation. (2) Pronuclei are formed. (3) The two pronuclei moving to the egg centre. (4) The two pronuclei apposing in the centre, just before PNMBD. (5) First cleavage furrow formed. (6) Nuclei form at the two-cell stage. White, blue and red

arrowheads indicate the 2pb, male and female pronucleus, respectively. In each frame, time is given in h:min after *in vitro* fertilization. Scale bars represent 50 μ m. **d**, Schematic models of first cleavage plane specification in each type of embryo (corresponding to the types in Fig. 1a–c and Fig. 2a–c).

specified only when the two pronuclei finally form the apposing plane in the centre of the egg just before, but not long before, entering M-phase. These experiments were performed using embryos from crosses of F1 (C57BL/6 X C3H) mice. To exclude possible strain influence we repeated the analysis using embryos from crosses of random-bred CD1 mice and obtained essentially the same results (data not shown).

Mouse centrosomes are derived from the oocyte and are distributed in the cytoplasm as a number of foci of small aster-like arrays of microtubules^{12,13}. Triple immunofluorescence staining of embryos from late interphase until the end of M-phase for micro-

tubules, actin and DNA was performed to examine their relationships and how they might bear on our model (Fig. 3). Consistent with previous reports^{12,13}, a few microtubule foci were associated with the periphery of the pronuclei, whereas many foci were still in the cytoplasm at late interphase (Fig. 3a); chromosomes started to condense, forming arrays around the nucleoli, and the microtubule foci were concentrated around the pronuclear membranes at the end of interphase (Fig. 3b). At prophase (Fig. 3c), both parental chromosome sets were surrounded by apparently irregular masses of microtubules and then formed a metaphase plate with a characteristic barrel-shaped spindle (Fig. 3d). At anaphase (Fig. 3e), the

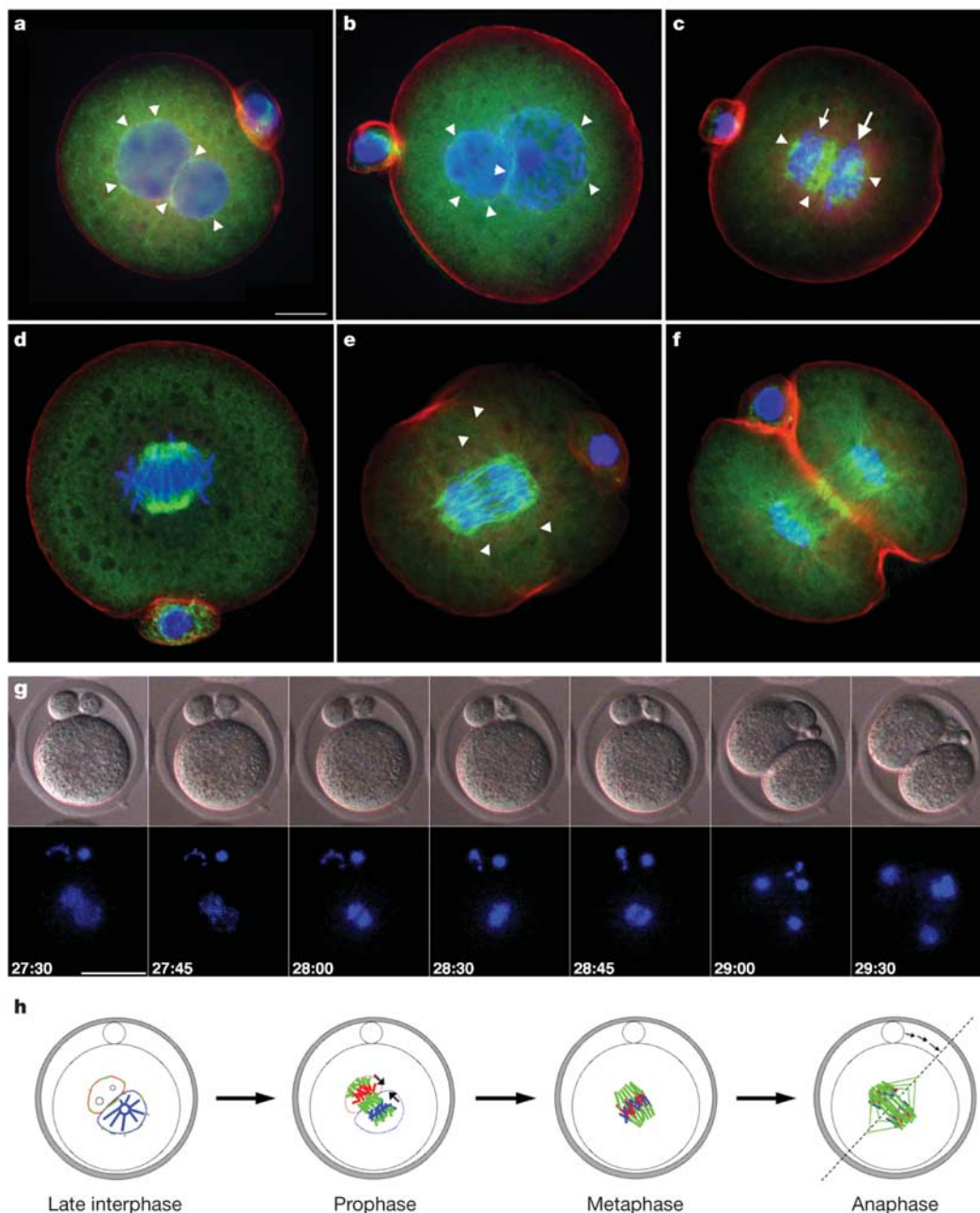


Figure 3 Relationship between cytoskeleton, chromosomes and first cleavage plane specification. **a–f**, Triple immunofluorescence staining of the embryo for microtubules (green), actin (red) and DNA (blue). Scale bar represents 20 μ m. **a**, Late interphase. **b**, The end of interphase. Arrowheads in **a** and **b** denote the small foci of microtubules. **c**, Prophase. Paternal and maternal chromosome aggregates (large and small arrow, respectively) are surrounded by masses of microtubules (arrowheads). **d**, Metaphase.

Metaphase plate with barrel-shaped spindle. **e**, Anaphase. Cleavage furrow is formed at the overlap of extended microtubules (arrowheads). **f**, Telophase. **g**, Time-lapse images of an embryo doubly recorded in DIC and fluorescence for DNA. Time is given in h:min after hCG injection. Scale bar represents 50 μ m. **h**, Schematic view of the process, summarized from **a–g**. Green, red and blue bars represent microtubules, maternal and paternal chromosomes, respectively.

first cleavage furrow formed at the overlap of extended microtubules at the cell cortex distal to the chromosomes¹⁴. To confirm these events chronologically, live embryos were stained with the fluorescent DNA dye Hoechst 33258, and both DIC and fluorescent time-lapse images were recorded (Fig. 3g). Note that the orientation of the movement of chromosomes from prophase to metaphase was parallel to that from metaphase to anaphase and telophase, enabling the embryo to use the same orientation of microtubule networks to effect the developmentally essential and unique mixing and separation of all the parental chromosomes (Fig. 3h).

To clarify whether the first cleavage plane is already determined in early interphase, or, as we propose here, is specified later according to the topology of the two apposing pronuclei, we conducted a series of experimental manipulations using embryos whose male pronucleus is formed adjacent to the female pronucleus (Fig. 4a, b). Based on our model, most of the embryos were expected to form the first cleavage plane passing close to the 2pb if allowed to develop undisturbed to the two-cell stage. The male pronucleus was removed from these embryos, and a male or female pronucleus from another embryo was transplanted to the opposite site of the enucleated zygote (Fig. 4c), generating a new topology between the two pronuclei. We then examined whether they follow their original information with respect to the site of division (blue dashed line in Fig. 4a), or instead form the first division plane based on the newly acquired topology (red lines in Fig. 4a). In the male pronucleus transfer experiments (Fig. 4d), 37 out of 45 embryos formed the cleavage plane according to the new topology of the pronuclei, with two having gradually rotated the apposing plane before finalizing the cleavage plane. In the female pronucleus transfers (Fig. 4e; and Supplementary Movie 6), 17 out of 18 embryos divided according to the new topology. Thus in 56 out of 63 manipulated embryos (89%) the cleavage plane was determined by the new topology of the pronuclei, in three (5%) the cleavage plane aligned with the 2pb position, presumably by chance, and in four (6%) the exact relationship with the plane of nuclear apposition could not be determined. As both the female and male pronuclei contributed similarly in defining the cleavage plane (Fig. 4d, e), a sperm component does not seem to be essential for this process. Furthermore preliminary experiments using parthenogenetically activated embryos showed that the first cleavage plane is defined by the two apposing pseudopronuclei in 18 out of 24 diploid parthenogenones (in six embryos the cleavage plane was not perpendicular to the observation plane, thus making a precise analysis of the relationship with the pseudopronuclei topology impossible) (data not shown). The cleavage plane in haploid parthenogenones containing a single pseudopronucleus was essentially random (data not shown).

Our present data show that the 2pb moves towards the first cleavage plane and does not serve as a stationary marker for the 'animal-pole' of the egg. Thus we conclude that there is no 'A-V axis' determined in the mouse egg. Furthermore, our data point to a mechanism by which the first cleavage plane is specified in the mouse egg, namely as the plane separating the two apposing pronuclei in the centre of the egg just before entering M-phase. This model contrasts sharply with previous claims^{5,6,8,9,15} that in 60% of embryos the site of the previous meiotic division (that is, the 'animal-pole') and the SEP⁸ define the first cleavage plane as passing preferentially within 30° of both these points. Our preliminary data indicate that sperm fuse with the oocyte preferentially (about 60%) in the half sphere overlying the meiotic chromosomes, except for the area directly above them¹⁶. Thus the previous claim^{5,6,8,9,15} could also be right. Almost half of the embryos were classified as type A (see Fig. 1a) and these are the embryos in which the cleavage plane passes within 30° on either side of the 2pb. This number is higher than expected, considering our preliminary data that about 30% of sperm bind to the arc of 60° centred on the 2pb. However, analysis of the position of the metaphase plates using immunofluorescently labelled embryos (Fig. 3d) demonstrated that the cleavage plane is

aligned within 30° of the 2pb in only 39% of embryos (12 out of 31). Thus the higher than expected proportion of type A embryos is most likely caused by the movement of the 2pb towards the expected cleavage plane even before cleavage begins.

When all these observations are taken into account it becomes apparent that the difference between the model presented here and those proposed previously^{5,6,8,9,15} is in the interpretation of the significance of the observed correlations. We claim that there is, as yet, no observable morphological feature in the unfertilized or just-fertilized mouse egg that can be used to predict with certainty the location of the first division plane. Once two pronuclei form, move to the centre and firmly appose, the cleavage plane will always coincide with the plane of apposition. Thus, our explanation as to what defines the plane of the first cleavage division is not based on statistical significance but is applicable to essentially all embryos. Furthermore, the analysis of cleavage plane formation in type B and C embryos (Fig. 2b, c), comprising half of the total embryos (Fig. 1), as well as in the embryos after the pronuclear transfer experiments (Fig. 4d, e) crucially test the validity of these opposing interpretations. According to the previously published models^{5,6,8,9,15}, the location of the cleavage plane in both of these situations would have been perpendicular to the one we predicted. However, the observed cleavage plane always corresponded to our prediction.

The molecular mechanisms underlying cleavage plane specification in the mouse egg remain obscure. The topological relationship between the two pronuclei must somehow be sensed and transmitted as a 'signal', possibly to the microtubule networks. The same polarity cue or microtubule networks might be used to bring the two parental chromosome sets to the centre, and to separate them

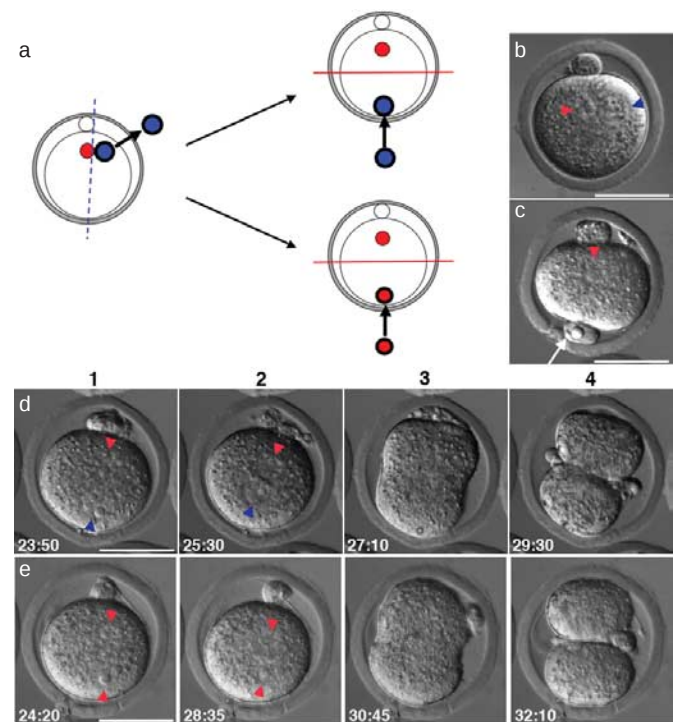


Figure 4 First cleavage plane is not determined in early interphase but is specified according to the newly formed topology of the pronuclei after manipulation. **a**, A schematic view of the design of the experimental manipulations. Male and female pronuclei are in blue and red, respectively. **b**, Zygote selected for the manipulation. **c**, Embryo immediately after transplantation of the pronucleus (white arrow) and before its fusion with the recipient enucleated zygote. **d, e**, Sequential images of the manipulated embryos after male (**d**) and female (**e**) pronucleus transfer. Red and blue arrowheads denote the female and the male pronuclei, respectively in **b–e**. Time is given in h:min after hCG injection. Scale bars represent 50 μ m.

equally¹⁷, a specific feature of the first cleavage shared by most animals. In the sea urchin egg, the same topological relationship might pertain¹⁸, however that notion is still somewhat controversial¹⁹. In the mouse egg, multiple centrosomes without centrioles²⁰ are distributed around the pronuclear membranes, and aggregate at both ends of the barrel-shaped spindle^{12,13}. It is not known whether centrosomes play a crucial role in transmitting a topological 'signal' to orient the microtubules, or whether microtubule networks organize according to the 'signal' from the chromosomes and thus navigate the centrosomes to further facilitate spindle formation^{21,22}.

The use of time-lapse recording technology provides a dynamic rather than static handle on mouse developmental events. This approach might also serve in revisiting the issue of the relationship of the first cleavage plane to the Em–Ab axis of the blastocyst and to the axis of bilateral symmetry of the gastrulating embryo, a question raised by our present reconstruction data and a recent study by others²³. □

Methods

Collection, *in vitro* culture and time-lapse recording of the embryos

For all experiments, MII oocytes or fertilized zygotes were recovered from superovulated (using human chorionic gonadotropin (hCG), 5 IU) (C57BL/6 X C3H) F1 female mice. These were mated with (C57BL/6 X DBA/2) F1 male mice in the case of zygotes (Figs 1, 3, 4), or *in vitro*-fertilized (Fig. 2). After brief treatment with hyaluronidase (300 U ml⁻¹), the embryos were washed with H-KSOM (KSOM with amino acids²⁴ and 21 mM Hepes) several times and cultured in a microdrop (2–3 µl) of H-KSOM mounted on glass chamber ZOG-3 (ELEKON science), covered with mineral oil (Sigma), in 5% CO₂ at 37°C. Temperature was maintained by Tempcontrol 37-2 digital (Zeiss) at 37.5°C on the indicator (it was confirmed by a microsensor that this kept the temperature of the recording drop at 36.9–37.0°C), together with a water bath connected to the heating stage (E100 with ecoline RE106, LAUDA) in a plastic chamber Incubator XL (Zeiss), attached to Zeiss Axiovert 200M with Narishige manipulators. Zeiss AxioVision Ver.3.1 software was used for the acquisition and analysis of the time-lapse images. The voltage of the halogen lamp was set below 2.6 V to minimize the embryo's exposure to the light. To observe the relationships of the two pronuclei, the 2pb and the cleavage plane, the manipulator was used to orient the embryo at the beginning (Figs 2–4, but not Fig. 1), so that the 2pb and the two pronuclei were all in the same focal plane. In most cases, the female pronucleus was formed very near the 2pb and the male pronucleus was formed just below the fertilization cone, so that initial orientation of the embryo according to the 2pb and the fertilization cone at the very beginning of the experiments in Fig. 2a–c practically obviated the need for fine-tuning after the pronuclei had formed (Fig. 2a–c sequences, between frames 2 and 3). However, some embryos could not be judged reliably because the cleavage plane was not close to perpendicular to the observation plane.

Methods for pronuclear transfer²⁵ and immunofluorescence staining are provided as Supplementary Methods

Received 18 December 2003; accepted 5 April 2004; doi:10.1038/nature02595.

1. Tarkowski, A. K. Experiments on the development of isolated blastomeres of mouse eggs. *Nature* **184**, 1286–1287 (1959).
2. Tarkowski, A. K. & Wroblewska, J. Development of blastomeres of mouse eggs isolated at the 4- and 8-cell stage. *J. Embryol. Exp. Morphol.* **18**, 155–180 (1967).
3. Rossant, J. Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J. Embryol. Exp. Morphol.* **36**, 283–290 (1976).
4. Tarkowski, A. K. Mouse chimaeras developed from fused eggs. *Nature* **190**, 857–860 (1961).
5. Gardner, R. L. The early blastocyst is bilaterally symmetrical and its axis of symmetry is aligned with the animal–vegetal axis of the zygote in the mouse. *Development* **124**, 289–301 (1997).
6. Gardner, R. L. Specification of embryonic axes begins before cleavage in normal mouse development. *Development* **128**, 839–847 (2001).
7. Piotrowska, K., Wianny, F., Pedersen, R. A. & Zernicka-Goetz, M. Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development. *Development* **128**, 3739–3748 (2001).
8. Piotrowska, K. & Zernicka-Goetz, M. Role for sperm in spatial patterning of the early mouse embryo. *Nature* **409**, 517–521 (2001).
9. Plusa, B., Grabarek, J. B., Piotrowska, K., Glover, D. M. & Zernicka-Goetz, M. Site of the previous meiotic division defines cleavage orientation in the mouse embryo. *Nature Cell Biol.* **4**, 811–815 (2002).
10. Schatten, G., Simerly, C. & Schatten, H. Microtubule configurations during fertilization, mitosis, and early development in the mouse and the requirement for egg microtubule-mediated motility during mammalian fertilization. *Proc. Natl Acad. Sci. USA* **82**, 4152–4156 (1985).
11. Maro, B., Johnson, M. H., Webb, M. & Flach, G. Mechanism of polar body formation in the mouse oocyte: an interaction between the chromosomes, the cytoskeleton and the plasma membrane. *J. Embryol. Exp. Morphol.* **92**, 11–32 (1986).
12. Maro, B., Howlett, S. K. & Webb, M. Non-spindle microtubule organizing centers in metaphase II-arrested mouse oocytes. *J. Cell Biol.* **101**, 1665–1672 (1985).
13. Schatten, H., Schatten, G., Mazia, D., Balczon, R. & Simerly, C. Behavior of centrosomes during fertilization and cell division in mouse oocytes and in sea urchin eggs. *Proc. Natl Acad. Sci. USA* **83**, 105–109 (1986).

14. Canman, J. C. *et al.* Determining the position of the cell division plane. *Nature* **424**, 1074–1078 (2003).
15. Piotrowska, K. & Zernicka-Goetz, M. Early patterning of the mouse embryo—contributions of sperm and egg. *Development* **129**, 5803–5813 (2002).
16. Johnson, M. H., Eager, D., Muggleton-Harris, A. & Grave, H. M. Mosaicism in organisation concanavalin A receptors on surface membrane of mouse egg. *Nature* **257**, 321–322 (1975).
17. Mayer, W., Smith, A., Fundele, R. & Haaf, T. Spatial separation of parental genomes in preimplantation mouse embryos. *J. Cell Biol.* **148**, 629–634 (2000).
18. Pawelcz, N., Mazia, D. & Finze, E. M. Fine structural studies of the bipolarization of the mitotic apparatus in the fertilized sea urchin egg. II. Bipolarization before the first mitosis. *Eur. J. Cell Biol.* **44**, 205–213 (1987).
19. Schatten, G. Sperm incorporation, the pronuclear migrations, and their relation to the establishment of the first embryonic axis: time-lapse video microscopy of the movements during fertilization of the sea urchin *Lytechinus variegatus*. *Dev. Biol.* **86**, 426–437 (1981).
20. Zamboni, L., Chakraborty, J. & Smith, D. M. First cleavage division of the mouse zygote. An ultrastructural study. *Biol. Reprod.* **7**, 170–193 (1972).
21. Heald, R. *et al.* Self-organization of microtubules into bipolar spindles around artificial chromosomes in *Xenopus* egg extracts. *Nature* **382**, 420–425 (1996).
22. Carazo-Salas, R. E. & Karsenti, E. Long-range communication between chromatin and microtubules in *Xenopus* egg extracts. *Curr. Biol.* **13**, 1728–1733 (2003).
23. Alarcon, V. B. & Marikawa, Y. Deviation of the blastocyst axis from the first cleavage plane does not affect the quality of mouse postimplantation development. *Biol. Reprod.* **69**, 1208–1212 (2003).
24. Ho, Y., Wigglesworth, K., Eppig, J. J. & Schultz, R. M. Preimplantation development of mouse embryos in KSOM: augmentation by amino acids and analysis of gene expression. *Mol. Reprod. Dev.* **41**, 232–238 (1995).
25. McGrath, J. & Solter, D. Nuclear transplantation in the mouse embryo by microsurgery and cell fusion. *Science* **220**, 1300–1302 (1983).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Y. Kaneda for HVJ; Z. Polanski, N. Bobola, A. Tomilin, P. Nielsen, R. Cassada and M. Hoffman for discussions and reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H. (hiiragi@immunbio.mpg.de).

Antero-posterior tissue polarity links mesoderm convergent extension to axial patterning

Hiromasa Ninomiya¹, Richard P. Elinson² & Rudolf Winklbauer¹

¹Department of Zoology, University of Toronto, Toronto, Ontario M5S 3G5, Canada

²Department of Biological Sciences, Duquesne University, Pittsburgh, Pennsylvania 15282, USA

Remodelling its shape, or morphogenesis, is a fundamental property of living tissue. It underlies much of embryonic development and numerous pathologies. Convergent extension (CE) of the axial mesoderm of vertebrates is an intensively studied model for morphogenetic processes that rely on cell rearrangement. It involves the intercalation of polarized cells perpendicular to the antero-posterior (AP) axis, which narrows and lengthens the tissue^{1,2}. Several genes have been identified that regulate cell behaviour underlying CE in zebrafish and *Xenopus*. Many of these are homologues of genes that control epithelial planar cell polarity in *Drosophila*^{1–5}. However, elongation of axial mesoderm must be also coordinated with the pattern of AP tissue specification to generate a normal larval morphology. At present, the long-range control that orients CE with respect to embryonic axes is not understood. Here we show that the chordamesoderm of *Xenopus* possesses an intrinsic AP polarity that is necessary for CE, functions in parallel to Wnt/planar cell polarity signalling, and determines the direction of tissue

elongation. The mechanism that establishes AP polarity involves graded activin-like signalling and directly links mesoderm AP patterning to CE.

To analyse the global control of CE, we dissociated chordamesoderm (prospective notochord) from the early gastrula and reaggregated cells after mixing to disturb the original spatial arrangement of cells (see Supplementary Fig. 1 for experimental design). We found that when prospective anterior and posterior halves of the chordamesoderm are labelled differently, partial sorting out is apparent at the neurula stage (Fig. 1b). At the tailbud stage sorting has progressed further, with anterior- and posterior-derived cells concentrated at the opposite ends of aggregates (Fig. 1d). This sorting indicates that the apparently homogenous chordamesoderm is patterned: cells express cues that allow them to associate with each other according to their original AP position.

Chordamesoderm AP patterning is also evident from the expression of *Xenopus Brachyury* (*Xbra*)⁶ and *chordin* (*chd*)⁷. In midgastrulae, *Xbra* expression is strongest at the blastopore lip and fades off towards the anterior chordamesoderm (Fig. 1h, left). The converse is true for *chd* expression (Fig. 1h, right). The *Xbra*–*chd* countergradient can be used to indicate the AP polarity of chordamesoderm explants. In Keller sandwich explants⁸, for example, expression of *Xbra* is strongest at one end (Fig. 1i, left), whereas that of *chd* peaks at the opposite end (Fig. 1i, right), indicating that the AP axis parallels that of elongation, as expected. The *Xbra*–*chd* countergradient, which like CE persists into neurula stages (data not shown), is independent evidence of chordamesoderm AP polarity. Whether *Xbra* or *chd* are determinants of this polarity, or only indicators, remains to be elucidated.

We found that AP polarity is linked to CE. At neurula stages, when Keller sandwiches have already elongated (Fig. 1i), explants containing cells from all AP levels in a mixed array show only irregular bulges (Fig. 1a). With progressive sorting elongation is resumed (Fig. 1c), however, suggesting that juxtaposition of anterior and posterior regions is required. To test this idea, anterior and posterior prospective chordamesoderm regions were

dissociated and reaggregated separately, and then recombined (Supplementary Fig. 1). We found that combinations of identical aggregates remain round (Fig. 1e, f), but combined anterior and posterior aggregates elongate (Fig. 1g; for quantitative data see Table 1). Shortly after combining different aggregates, when explants are still flat from centrifugation, a distinct gene expression boundary reflects cell lineage (Fig. 1j), but at neurula stages expression has become graded (Fig. 1k). This implies that AP polarity is restored, triggering CE. As a corollary, sorting in mixed explants is expected to be vigorous enough to outcompete the simultaneous levelling of positional values.

Activin-treated animal caps are commonly used to study CE. We found that activin-induced chordamesoderm formation is not sufficient for CE; in addition, AP polarity must be established. This can be achieved through treatment with graded doses of activin (Fig. 2a). In our system, uniform exposure of dissociated cap cells, pretreated with LiCl⁹, to 0.5 or 1 ng ml^{−1} of activin induces notochord ('uniform explants'). By contrast, reaggregates in which superficial cells are better exposed to activin than are deeper cells, require 5 ng ml^{−1} of activin; this yields notochord-containing explants that are 'graded' with respect to activin exposure (Fig. 2b, c). Graded explants elongate (Fig. 2f), whereas uniform explants remain spherical (Fig. 2g and Table 1). In addition, in uniform explants *chd* and *Xbra* are evenly distributed (Fig. 2o), with higher doses of activin yielding more anterior-type gene expression (data not shown). In graded explants, a *chd*–*Xbra* countergradient (Fig. 2n) indicates an AP axis coincident with the axis of elongation. Thus, differential activin induction may establish AP polarity and CE.

To test this notion, we combined aggregates treated uniformly with high or low activin concentrations (Fig. 2a). We found that combining identically treated aggregates yields globular explants (Fig. 2h–j), whereas elongation occurs when cells treated with low and medium (Fig. 2k) or with medium and high doses (Fig. 2l) are combined. Thus, it is not a specific activin dose that is required, but rather the interaction of cells exposed to different doses. Mixing these cells homogeneously does not promote elongation, however (Fig. 2m and Table 1), although both combined and mixed explants form notochord (Fig. 2d, e). Apparently, the parallel interaction of many cells is required, as provided at the interface between populations, probably to re-establish AP polarity. Indeed, gene expression gradients are restored in explants (Fig. 2p, q). Thus, different doses of activin induce distinct AP positional values in

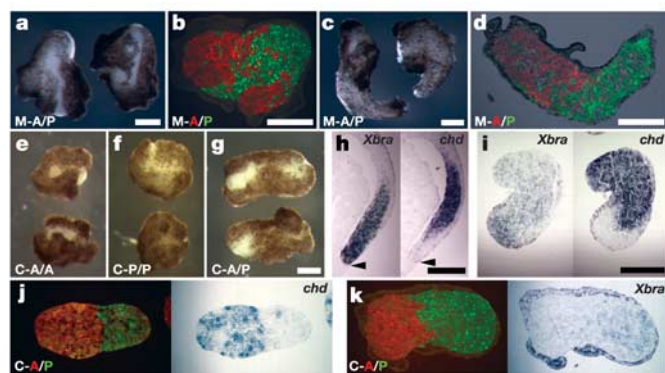


Figure 1 AP polarity and chordamesoderm explant elongation. **a–d**, Mixed explants at the neurula (**a, b**) and tailbud (**c, d**) stage, viewed externally (**a, c**) and in sections (**b, d**). Prospective posterior chordamesoderm cells (animal one-third of dorsal lip) are green, anterior cells (middle one-third of lip) are red. **e–g**, Combined explants at the neurula stage. Shown are combinations of two anterior (**e**), two posterior (**f**), and one anterior and one posterior (**g**) aggregates. **h, i**, Section *in situ* hybridization of a stage-12 gastrula, dorsal side (**h**), and a stage-15 Keller sandwich (**i**) explant, showing expression of *Xbra* (left images) and *chd* (right images) in the chordamesoderm. Anterior is to the top; arrowheads indicate the blastopore. **j, k**, Combined explants after 1 h (**j**) and at the neurula stage (**k**). Cell lineage in **j** and **k** is the same as in **b** and **d**, respectively, with expression of *chd* (**j**, right) and *Xbra* (**k**, right) visualized by section *in situ* hybridization. Explants are initially flat from centrifugation (**j**), round up, and then elongate (**k**). See footnotes to Table 1 for abbreviations. Scale bars, 200 μ m.

Table 1 Elongation of explants

Type of explant†	Elongation*			
	None	Moderate	Extensive	Moderate/extensive (%)
Chordamesoderm explants				
Keller sandwich	4	9	9	82
M-A/P (neurula)	8	6	0	43
M-A/P (tailbud)	0	1	6	100
C-A/A or P/P	24	3	1	14
C-A/P	0	13	1	100
Activin-induced explants‡				
G-5	4	7	7	78
U-1	16	0	0	0
C-0.2/0.2, 0.5/0.5, 1/1	25	1	0	4
C-0.2/0.5, 0.5/1	3	23	6	91
M-0.5/1	10	3	0	23
Activin induced explants: effect of dnXdsh‡				
G-5 (β)	2	4	11	88
G-5 (Δ)	15	3	1	21
C-0.5/1 (β/β)	3	3	6	75
C-0.5/1 (Δ/Δ)	11	1	0	8

*Elongation was evaluated by determining the length-to-width (L/W) ratio: none, L/W < 2; moderate elongation, L/W > 2; extensive elongation, L/W > 3.

†M, mixed explant; A, anterior; P, posterior; C, combined explant; G, graded explant; U, uniform explant; β , β -galactosidase; Δ , *Xdishevelled* mutant Δ PDZ; dnXdsh, dominant-negative Xdsh.

‡Numbers after hyphen indicate the concentration of activin (in ng ml^{−1}).

uniform explants, but interaction between cell populations re-establishes the graded pattern that is correlated with CE. In induced intact animal caps, an intrinsic heterogeneity probably supports AP patterning and CE: caps are prepatterned, with different regions responding differently to induction^{10,11}, and cells are unequally exposed to activin¹².

Our findings suggest that the same mechanism that generates the AP sequence of dorsal mesodermal tissues also orients CE by establishing AP polarity in a subdomain of this mesoderm. *In vitro* activin acts as a morphogen, with low concentrations inducing *Xbra*-expressing axial and/or paraxial mesoderm, and high concentrations inducing *gsc*-expressing prechordal mesoderm^{13,14}. Most significantly, activin can establish graded differences in gene expression at the single cell level¹⁵. Although activin is conveniently used *in vitro*, nodal-related factors may function as the main endogenous ligands, and experimental manipulation of ligand activity in *Xenopus* and zebrafish embryos suggests that graded nodal signalling controls AP patterning^{16–18}. Our data complement these findings by showing that within a single tissue—the chordamesoderm—differential activin-like signalling yields different AP positional values. Again, lower activin concentrations induce more posterior, and higher ones more anterior, positions. The dependence of mesodermal AP tissue specification, as well as AP polarity implementation, on a common mechanism elegantly ensures par-

allel alignment of the AP axis and CE. Germband extension along the AP axis in *Drosophila* uses the same strategy of control: it depends on AP and not on mediolateral patterning of the blastoderm. In this case, however, a periodic pattern of alternating stripes, defined by pair-rule gene expression, is required instead of graded positional values^{19,20}. It will be interesting to see whether common molecular mechanisms operating downstream of AP patterning are involved in these two systems.

We determined that explant constriction and elongation at the boundary between differentially induced populations are accompanied by cell intercalation, by combining four labelled aggregates: two incubated at low and two incubated at high activin concentrations (Fig. 3a). We found that after elongation cells in the constricted zone are aligned perpendicular to the axis of elongation and have intercalated (Fig. 3b), with their direction of movement being determined by AP polarity. To confirm that the induced elongation represents CE, we interfered with the Wnt/planar cell polarity (PCP) pathway by expressing a *Xenopus Dishevelled* (*Xdsh*) construct that specifically blocks non-canonical Wnt signalling^{21–23}.

We found that graded explants expressing $\Delta PDZ^{22,23}$ elongate less than controls (Fig. 3c, d). ΔPDZ also impedes elongation in

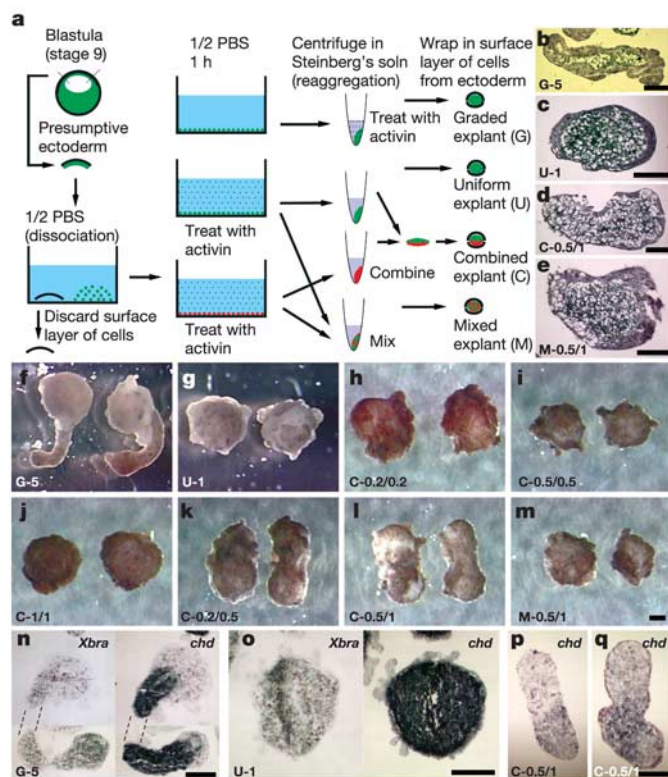


Figure 2 Activin-induced elongation. **a**, Explant preparation. **b–e**, Sections of tailbud stage explants stained with MZ15 antibody, showing vacuolated notochord cells in graded (**b**), uniform (**c**), combined (**d**) and mixed (**e**) explants. **f–m**, Morphology of graded (**f**), uniform (**g**), combined (**h–j**) and mixed (**m**) explants at the neurula stage. Each of the two aggregates combined was either induced identically (**h–j**) or differently (**k, l**). **n–q**, *In situ* hybridization of explant sections. **n, o**, Expression of *Xbra* (left images) and *chd* (right images) in graded (**n**) and uniform (**o**) explants at the neurula stage. The dotted lines (**n**) indicate the connecting part between anterior (upper) and posterior (lower) portions outside the plane of section. **p, q**, Expression of *chd* in combined explants after 1 h (**p**) and at the neurula stage (**q**). The discrete expression boundary (**p**) corresponds to lineage boundary (not shown). See footnotes to Table 1 for abbreviations. Scale bars, 200 μ m.

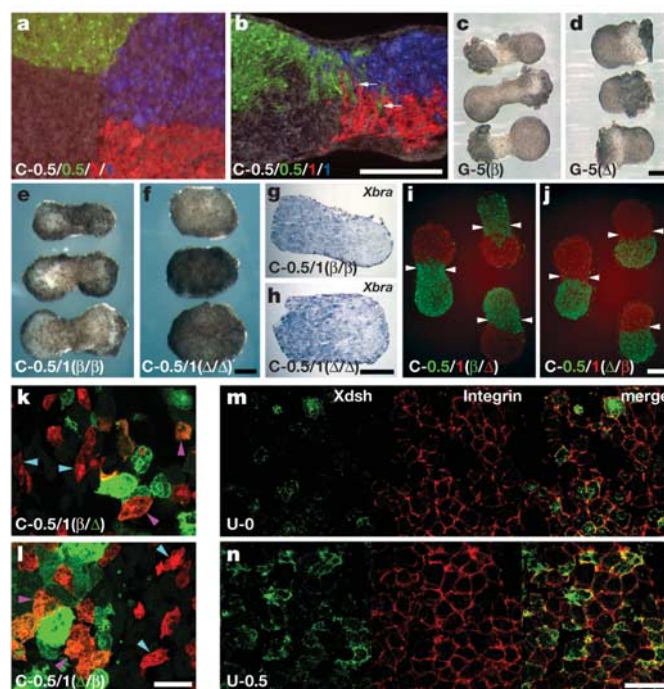


Figure 3 Convergent extension and Wnt/PCP signalling. **a, b**, Cell intercalation 1 h after combining aggregates (**a**), and at stage 25 (**b**). Aggregates were either non-labelled or labelled with fluorescein dextran amine (FDA; green), rhodamine dextran amine (RDA; red) and Cascade Blue dextran amine (CBDA; blue). Arrows in **b** indicate intercalating cells. **c, d**, Graded explants injected with β -galactosidase (β -gal) mRNA (**c**) or ΔPDZ (**d**) mRNA. **e–h**, Combined explants injected with β -gal (**e**) or ΔPDZ (**f**) mRNA, and *Xbra* expression in β -gal- (**g**) and ΔPDZ -injected (**h**) explants. **i–l**, ΔPDZ is expressed in one of the two combined aggregates, whereas β -gal is expressed in the other. Maximal constriction (white arrowheads) is in the β -gal-expressing part, regardless of whether it received the lower (**i**) or higher (**j**) activin dose. **k, l**, RDA labelling (red) of a fifth of the cells in both parts shows that cells on the β -gal side appear polarized and aligned (blue arrowheads), whereas cells on the ΔPDZ side (green, ΔPDZ -GFP) do not (purple arrowheads), regardless of whether the activin dose was higher (**k**) or lower (**l**) on the ΔPDZ side. **m, n**, *Xdsh*-GFP localization in untreated (**m**) and activin-treated (**n**) uniform explants. Left, *Xdsh*-GFP; middle, 8C8 antibody against integrin, demarcating cell membranes; right, overlay of *Xdsh* (green) and 8C8 (red) signal. See footnotes to Table 1 for abbreviations. Scale bars, 200 μ m, except for those in panels **i** and **n** (50 μ m).

combined explants (Fig. 3e, f and Table 1) without compromising graded gene expression (Fig. 3g, h). When one of the combined aggregates expresses ΔPDZ , only the control-injected portion elongates and constriction is always in this part (Fig. 3i, j). In addition, cells become elongated and aligned in the control-injected part, but are irregularly shaped in the ΔPDZ -expressing region (Fig. 3k, l). Thus, *Xdsh* controls elongation autonomously and is not required for establishing AP polarity. Independent elongation of the two halves of explants argues that it is not due to the separation of differentially adhesive cell populations²⁴. Its dependence on *Xdsh* also confirms that this process represents CE. Because impairing Wnt/PCP signalling does not seem to prevent establishment of AP polarity, the latter should function upstream or in parallel to the PCP pathway.

An indicator of Wnt/PCP signalling associated with CE is the membrane recruitment of *Xdsh* protein^{23,25}. We found that uniform activin treatment is sufficient for this relocalization. Whereas untreated reaggregates show a punctate distribution of cytoplasmic *Xdsh* (Fig. 3m), this molecule is enriched at cell boundaries in aggregates of uniformly induced cells (Fig. 3n). Thus, expression of AP polarity is not required for *Xdsh* translocation to the membrane, and the failure of uniform explants to elongate is probably not due to a lack of PCP signalling. Instead, AP polarity seems to be an additional requirement for CE.

Two features characterize the cellular basis of amphibian CE: cells are 'bipolar', with locomotory processes at opposite ends; and cells are all 'oriented' according to a global pattern, namely, perpendicular to the AP axis. It is not clear whether these two components are independently controlled or are the expression of a single underlying mechanism. In addition, PCP signalling is necessary for this specific arrangement of cells^{23,25,26}, but it has not been strictly determined whether it controls cell polarity, cell orientation or both, or whether it is instructive or permissive. In zebrafish, the *Wnt11*-deficient *silberblick*^{-/-} phenotype is rescued by injecting *Wnt11* or ΔN -*dsh* messenger RNA⁵, and in *Xenopus* the effect of dominant-negative *Wnt11* can be reversed by co-injecting *Xdsh* mRNA⁴. This suggests that Wnt/PCP signalling may be permissive for cell alignment. Moreover, in zebrafish *trilobite/strabismus* PCP mutants, dorsal migration of lateral cells during CE is attenuated, but the orientation of movement is not affected²⁷, implying that the latter is controlled independently of the PCP pathway.

Our data indicate that it may be AP polarity that orients cells relative to the body axes. Its function could be viewed as analogous to the long-range signal that has been proposed to function in the developing *Drosophila* wing or eye, namely to bring cells that are polarized by PCP signalling in alignment with the body axes²⁸. Alternatively, AP polarity could control cell polarity and orientation simultaneously, analogous to the mechanism proposed for *Drosophila* germband extension²⁰. With protrusive activity being suppressed on anterior and posterior cell surfaces, cells would be functionally bipolar and properly oriented. PCP signalling would be an essential, yet permissive, component of this model. □

Methods

Preparation of explants

Explants were prepared as shown in Fig. 2a and Supplementary Fig. 1, and as described in the Supplementary Methods. Explants or embryos were fixed in MEMFA and prepared for Paraplast sectioning, to evaluate lineages, *in situ* hybridization or notochord differentiation. We distributed consecutive sections over several slides together with sections from control embryos.

Analysis of explants

Notochord differentiation was detected by using antibody MZ15 (ref. 29). *In situ* hybridization on sections was done as described on the De Robertis' Laboratory Home Page (<http://www.lifesci.ucla.edu/hhmi/derobertis>).

To detect green fluorescent protein (GFP)-conjugated *Xdsh* constructs (wild type and $\Delta PDZ/D2$), we fixed explants in 4% paraformaldehyde in PBS and fractured, mounted and observed them under a confocal microscope. To visualize cell membranes, explants were counterstained with monoclonal antibody 8C8 against *Xenopus* integrin- β A (ref. 30) after fracturing.

Received 9 February; accepted 4 May 2004; doi:10.1038/nature02620.

- Keller, R. Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* **298**, 1950–1954 (2002).
- Wallingford, J. B., Fraser, S. E. & Harland, R. M. Convergent extension: the molecular control of polarized cell movement during embryonic development. *Dev. Cell* **2**, 695–706 (2002).
- Mlodzik, M. Planar cell polarization: do the same mechanisms regulate *Drosophila* tissue polarity and vertebrate gastrulation? *Trends Genet.* **18**, 564–571 (2002).
- Tada, M. & Smith, J. C. *Xwnt11* is a target of *Xenopus* Brachyury: regulation of gastrulation movements via dishevelled, but not through the canonical Wnt pathway. *Development* **127**, 2227–2238 (2000).
- Heisenberg, C. P. *et al.* Silberblick/*Wnt11* mediates convergent extension movements during zebrafish gastrulation. *Nature* **405**, 76–81 (2000).
- Smith, J. C., Price, B. M., Green, J. B., Weigel, D. & Herrmann, B. G. Expression of a *Xenopus* homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67**, 79–87 (1991).
- Sasai, Y. *et al.* *Xenopus* chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell* **79**, 779–790 (1994).
- Keller, R. & Danilchik, M. Regional expression, pattern and timing of convergence and extension during gastrulation of *Xenopus laevis*. *Development* **103**, 193–209 (1988).
- Klein, P. S. & Melton, D. A. A molecular mechanism for the effect of lithium on development. *Proc. Natl Acad. Sci. USA* **93**, 8455–8459 (1996).
- Sokol, S. & Melton, D. A. Pre-existent pattern in *Xenopus* animal pole cells revealed by induction with activin. *Nature* **351**, 409–411 (1991).
- Ninomiya, H., Takahashi, S., Tanegashima, K., Yokota, C. & Asashima, M. Endoderm differentiation and inductive effect of activin-treated ectoderm in *Xenopus*. *Dev. Growth Differ.* **41**, 391–400 (1999).
- McDowell, N., Zorn, A. M., Crease, D. J. & Gurdon, J. B. Activin has direct long-range signalling activity and can form a concentration gradient by diffusion. *Curr. Biol.* **7**, 671–681 (1997).
- Green, J. B., New, H. B. & Smith, J. C. Responses of embryonic *Xenopus* cells to activin and FGF are separated by multiple dose thresholds and correspond to distinct axes of the mesoderm. *Cell* **71**, 731–739 (1992).
- Gurdon, J. B., Harger, P., Mitchell, A. & Lemaire, P. Activin signalling and response to a morphogen gradient. *Nature* **371**, 487–492 (1994).
- Gurdon, J. B. *et al.* Single cells can sense their position in a morphogen gradient. *Development* **126**, 5309–5317 (1999).
- Thisse, B., Wright, C. V. E. & Thisse, C. Activin- and nodal-related factors control antero-posterior patterning of the zebrafish embryo. *Nature* **403**, 425–428 (2000).
- Gritsman, K., Talbot, W. S. & Schier, A. F. Nodal signaling patterns the organizer. *Development* **127**, 921–932 (2000).
- Branford, W. W. & Yost, H. J. Lefty-dependent inhibition of Nodal- and Wnt-responsive organizer gene expression is essential for normal gastrulation. *Curr. Biol.* **12**, 2136–2141 (2002).
- Irvine, K. D. & Wieschaus, E. Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development* **120**, 827–841 (1994).
- Zallen, J. A. & Wieschaus, E. Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell* **6**, 343–355 (2004).
- Sokol, S. Y., Klingensmith, J., Perrimon, N. & Itoh, K. Dorsalizing and neuralizing properties of *Xdsh*, a maternally expressed *Xenopus* homolog of dishevelled. *Development* **121**, 1637–1647 (1995).
- Rothbacher, U. *et al.* Dishevelled phosphorylation, subcellular localization and multimerization regulate its role in early embryogenesis. *EMBO J.* **19**, 1010–1022 (2000).
- Wallingford, J. B. *et al.* Dishevelled controls cell polarity during *Xenopus* gastrulation. *Nature* **405**, 81–85 (2000).
- Townes, P. L. & Holtfreter, J. Directed movements and selective adhesion of embryonic amphibian cells. *J. Exp. Zool.* **128**, 53–120 (1955).
- Kinoshita, N., Iio, H., Miyakoshi, A. & Ueno, N. PKC δ is essential for Dishevelled function in a noncanonical Wnt pathway that regulates *Xenopus* convergent extension movements. *Genes Dev.* **17**, 1663–1676 (2003).
- Goto, T. & Keller, R. The planar cell polarity gene *strabismus* regulates convergence and extension and neural fold closure in *Xenopus*. *Dev. Biol.* **247**, 165–181 (2002).
- Jessen, J. R. *et al.* Zebrafish *trilobite* identifies new roles for *strabismus* in gastrulation and neuronal movements. *Nature Cell Biol.* **4**, 610–615 (2002).
- Strutt, D. Frizzled signalling and cell polarisation in *Drosophila* and vertebrates. *Development* **130**, 4501–4513 (2003).
- Smith, J. C. & Watt, F. M. Biochemical specificity of *Xenopus* notochord. *Differentiation* **29**, 109–115 (1985).
- Gwantka, V., Ellinger-Ziegelbauer, H. & Hausen, P. β 1-Integrin is a maternal protein that is inserted into all newly formed plasma membranes during early *Xenopus* embryogenesis. *Development* **115**, 595–605 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the National Hormone & Pituitary Program and A. F. Parlow for recombinant human activin A; F. Watt for MZ15 antibody; P. Hausen for 8C8 antibody; E. M. DeRobertis, R. Harland and J. Smith for plasmids; Y. Masui and the members of the Elinson and Winklbauer laboratories for help and encouragement; and M. Makowiecki for manuscript suggestions. This work was supported by an International Collaborative Grant from the Human Frontier Science Program Organization to R.P.E. and by grants from the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, and the Canada Foundation for Innovation to R.W.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W. (winklbauer@zoo.utoronto.ca).

Dynamic control of positional information in the early *Drosophila* embryo

Johannes Jaeger¹, Svetlana Surkova², Maxim Blagov², Hilde Janssens¹, David Kosman³, Konstantin N. Kozlov², Manu¹, Ekaterina Myasnikova², Carlos E. Vanario-Alonso^{1,4}, Maria Samsonova², David H. Sharp⁵ & John Reintz¹

¹Department of Applied Mathematics and Statistics, and Center for Developmental Genetics, Stony Brook University, Stony Brook, New York 11794-3600, USA

²Department of Computational Biology, Center for Advanced Studies, St Petersburg State Polytechnic University, St Petersburg, 195251 Russia

³Department of Biology, University of California, San Diego, California 92093, USA

⁴Universidade Federal do Rio de Janeiro, Instituto de Biofísica Carlos Chagas Filho, Rio de Janeiro, Rio de Janeiro 21949-900, Brazil

⁵Applied Physics Division, and Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

Morphogen gradients contribute to pattern formation by determining positional information in morphogenetic fields^{1,2}. Interpretation of positional information is thought to rely on direct, concentration-threshold-dependent mechanisms for establishing multiple differential domains of target gene expression^{1,3,4}. In *Drosophila*, maternal gradients establish the initial position of boundaries for zygotic gap gene expression, which in turn convey positional information to pair-rule and segment-polarity genes, the latter forming a segmental pre-pattern by the onset of gastrulation^{5–7}. Here we report, on the basis of quantitative gene expression data, substantial anterior shifts in the position of gap domains after their initial establishment. Using a data-driven mathematical modelling approach^{8–11}, we show that these shifts are based on a regulatory mechanism that relies on asymmetric gap–gap cross-repression and does not require the diffusion of gap proteins. Our analysis implies that the threshold-dependent interpretation of maternal morphogen concentration is not sufficient to determine shifting gap domain boundary positions, and suggests that establishing and interpreting positional information are not independent processes in the *Drosophila* blastoderm.

The maternal Bicoid (Bcd) gradient in blastoderm-stage embryos of *Drosophila melanogaster* is a classic example of a morphogen gradient^{12,13}. The modern morphogen concept⁴ implies that Bcd by itself should be able to specify directly the domain boundaries of its target genes, such as the gap genes *Krüppel* (*Kr*), *giant* (*gt*) (Fig. 1a), *knirps* (*kni*) and *hunchback* (*hb*) (Fig. 1b). Genetic and theoretical studies suggest, however, that gap gene regulation by Bcd requires regulatory synergism with maternal Hb^{9,14}, and quantitative experiments show that Bcd by itself cannot account for the precise positioning of zygotic *hb* expression^{12,15}. In addition, sharpening of gap domain boundaries and maintenance of gap gene expression rely on gap–gap cross-repression (see ref. 16, and references therein). It remains unclear, however, whether such cross-repression is required for correct positioning of gap gene boundaries, or if a combination of the maternal Bcd, Hb and Caudal (Cad) gradients is sufficient to determine positional information in the gap gene system.

Proof of the sufficiency of a given set of genetic interactions for specific expression patterns can be achieved only by reconstituting the underlying gene network from well-defined ingredients. Because this is currently impossible *in vitro*, we have used an *in silico* approach. The gene circuit method is a data-driven mathematical modelling approach for the computational reconstitution and

Box 1

The four steps of the gene circuit method

Step 1: Formulation of the model

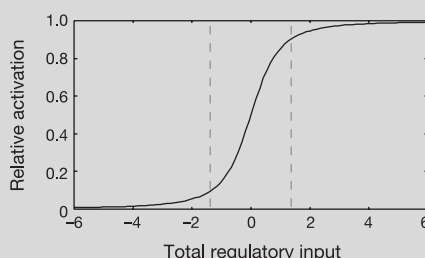
Gene circuits are hybrid dynamical models consisting of discrete nuclear divisions and continuous intranuclear dynamics of protein concentrations. Nuclei indexed by *i* are arranged in a one-dimensional row along the A–P axis, because A–P and D–V patterning systems are largely independent in the trunk region of the embryo. Gap gene circuit models cover 58 nuclei between 35 and 92% A–P position during cycles 13 and 14A²⁷, and include *bcd*, *cad*, *hb*, *Kr*, *kni*, *gt* and the terminal gap gene *tl*. The rates of change in protein concentration dv_i^a/dt for each regulated gene product *a* in each nucleus *i* during interphase are given by a system of 348 ordinary differential equations defined by

$$\frac{dv_i^a}{dt} = R_a g(u^a) + D^a [(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)] - \lambda_a v_i^a \quad (1)$$

The main terms on the right hand side of equation (1) represent protein synthesis, diffusion and decay respectively, with the corresponding rate parameters R_a , D^a and λ_a . Diffusion parameters are proportional to the inverse square distance between nuclei, which is halved on division. During mitosis, protein synthesis is shut down. Nuclear division occurs at the end of mitosis in cycle 13. $g(u^a)$ is a sigmoidal regulation–expression function, where u^a is given by

$$u^a = \sum_b T^{ab} v_i^b + m^a v_i^{\text{Bcd}} + h^a \quad (2)$$

For values of u^a below -1.5 and above $+1.5$, $g(u^a)$ rapidly approaches zero and one, respectively (broken lines in figure below).



Parameters T^{ab} constitute a genetic interconnectivity matrix and represent activation of gene *a* by the product of gene *b* if positive, repression if negative, and no interaction if close to zero. v_i^{Bcd} represents the concentration of Bcd in nucleus *i*, which is exclusively maternal and constant in time. m^a describes the regulatory input of Bcd to the zygotic system. Maternal contributions to Hb and Cad are represented as nonzero initial conditions. h^a represents regulatory input from ubiquitous maternal factors (see Supplementary Information for details).

Step 2: Quantitative gene expression data

Quantified expression profiles for products of genes *a* at one time point in cycle 13 and eight time points during cycle 14A are used for fits of the model to data (see Methods).

Step 3: Optimization to fit model to data

Parameter values are not fixed *a priori*, but are obtained by fitting the model to data. Numerically calculated model output is compared with quantitative gene expression data, and the difference between the two is minimized by adjusting parameter values using the PLSA optimization method (see Methods). In this way, we obtain specific sets of parameters, and thus specific gene circuits, which faithfully reproduce the expression patterns used for optimization (see Supplementary Information for parameter values).

Step 4: Biological analysis

Regulatory parameters of gene circuits contain information about the generative mechanism underlying the expression data used for optimization. We extract this information by graphical analysis of both the reaction and diffusion terms of equation (1), and the quantitative regulatory contributions $T^{ab} v_i^b$ by regulators *b* to specific genes *a* in each nucleus *i* in equation (2).

analysis of observed gene expression patterns, which allows us to infer regulatory interactions from wild-type gene expression data^{8–11,16} (Box 1).

Quantitative gene expression data show significant anterior shifts of gap domain boundaries during cleavage cycle 14A (Fig. 1c, e, g). Gap gene circuit models reproduce observed gap gene expression patterns, including boundary shifts of the central *Kr* domain and the posterior domains of *kni* and *gt*, with high precision and temporal resolution (Fig. 1d, f, h). Elsewhere, we have shown that our models also specifically and accurately reproduce the activation of gap genes by maternal factors and gap–gap cross-repression¹⁶. Taken together, this suggests that gap gene circuit models correctly represent the dynamic and genetic properties of the gap gene system.

In gap gene circuits, dynamic shifts of gap gene expression domains are reflected at the level of the rate of change in protein concentration (Fig. 2a–c). We can distinguish two discrete regulatory domains in the anterior and posterior portions of the *Kr*, *kni* and *gt* expression domains: protein synthesis dominates anteriorly, whereas protein decay posteriorly (Fig. 2d–f). The combination of anterior synthesis and posterior decay leads to an anterior shift of each expression domain. In addition, both the synthesis and decay domains themselves shift anteriorly over time (Fig. 2a–c).

The model thus predicts that synthesis is confined to the anterior region of each expression domain, implying that there is an asymmetric distribution of gap gene transcript in protein domains. Confirmation of this prediction is shown in Fig. 3. During cycle 14A, transcript domains of *Kr*, *kni* and *gt* are shifted anteriorly with respect to their corresponding protein domains (Fig. 3e–q). Such an asymmetrical distribution of *Kr* transcript and protein domains has been reported previously¹⁷, but it has not been interpreted in terms of dynamical shifts. In addition, anterior shifts of domain boundaries are present at both the transcript and the protein level (Fig. 3),

consistent with shifts in domains of synthesis, as observed in gene circuits (Fig. 2a–c).

A generative mechanism for gap domain shifts must explain the dynamic positioning of the domains of protein production and decay. Theoretically, shifts in domain boundaries can be caused by mechanisms based on gene regulatory interactions and/or the diffusion of gene products between neighbouring nuclei. In the syncytial blastoderm of *Drosophila*, the absence of cell membranes between nuclei allows the diffusion of gap gene products, leading to protein domains that, in general, are slightly larger than their corresponding transcript domains (Fig. 3).

Gene circuits allow us to examine the roles of regulatory and diffusive mechanisms by plotting the corresponding terms in the gene circuit equation (Box 1). Here, we illustrate our analysis with the example of the posterior boundary of *kni*. We compare temporal changes in protein concentration (Fig. 4a–c) to temporal changes in the reaction (protein synthesis and decay) and the diffusion terms of the equation (Fig. 4e–g). Nuclei that lie in the zone of the anterior shift of the posterior *kni* boundary show a characteristic switch from synthesis to decay of *Kni* protein during cycle 14A (Fig. 4a–c). This dynamical behaviour is directly responsible for the observed anterior shift of the boundary. It is closely mirrored by changes in the reaction terms of the equation, whereas diffusion leads to an influx of protein into the shift zone counteracting the boundary shift (Fig. 4e–g). For shifts in posterior borders, a counteracting role of diffusion is expected because the shift occurs toward increasing protein concentration in the boundary. But even for shifts in anterior boundaries, which occur in the direction of lower concentration, diffusion is not essential because these shifts are still present in gap gene circuits in which diffusion is not allowed to occur (see Supplementary Information). Thus, our analysis suggests that although gap protein diffusion is present in both embryo and gap gene circuits, it does not have a significant role in shifting gap domain boundaries.

Gene circuits allow us to study gene regulatory interactions by plotting combinations of regulatory contributions to the expression of a particular gene. For the shift in the posterior boundary of *kni*, we have found the following relevant regulatory contributions (Fig. 4i–k). Expression of *kni* is repressed in nuclei in the shift zone by spatially specific repressive inputs that counteract broad *Cad* activation throughout the posterior region of the embryo (Fig. 4i–k). Downregulation of *kni* in the posterior portion of the

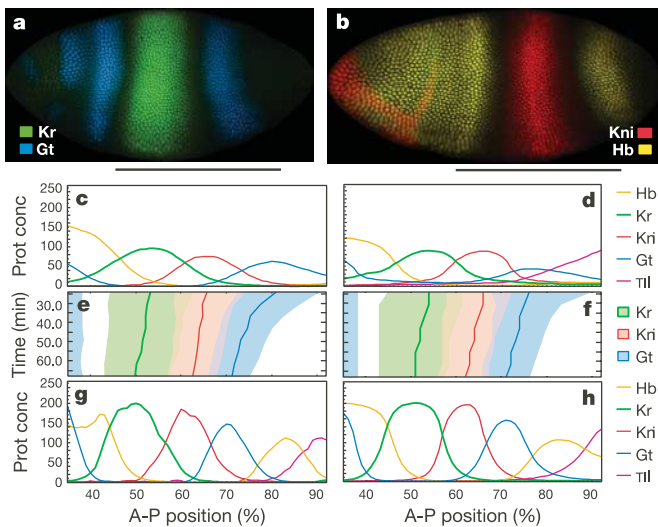


Figure 1 Dynamical shifts in gap gene domains are reproduced by gap gene circuits. **a, b**, *Drosophila melanogaster* blastoderm stage embryos at late cleavage cycle 14A (time class T8) immunostained for *Kr* and *Gt* (FlyEx embryo, *rge9*; **a**) and *Kni* and *Hb* (*rb8*; **b**). Anterior is to the left, dorsal is up. Bars indicate the region included in gap gene circuits. **c, d, g, h**, Gene expression data (**c, g**) and gap gene circuit model output (**d, h**) at early (T1; **c, d**) and late (T8; **g, h**) cycle 14A. Vertical axes represent relative protein concentrations, horizontal axes represent position along the A–P axis (where 0% is the anterior pole). There are no T1 data for T1 (**c**). **e, f**, Gap domain shifts for *Kr*, *kni* and *gt* covering the time between patterns shown in **c, d** and **g, h**. Lines indicate the position of maximum concentration for each domain. Coloured areas represent regions in which protein concentration is above the half-maximum value. Positional values for data were obtained by approximation with quadratic splines²².

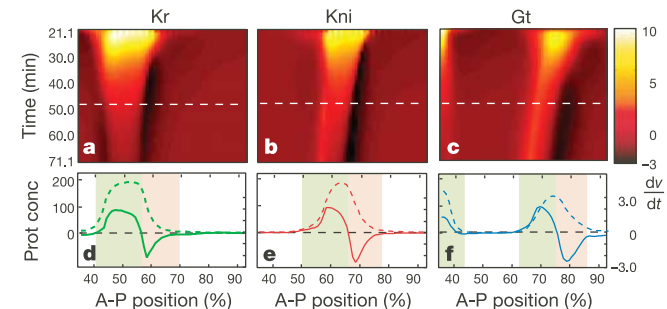


Figure 2 Shifting domains of gap protein synthesis and decay. **a–c**, Time–space diagrams, based on model output, of rate of change in protein concentration, dv/dt , for *Kr* (**a**), *Kni* (**b**) and *Gt* (**c**) during cycle 14A. Vertical axes represent time, horizontal axes represent position along the A–P axis. Note the shifting positions of domains of protein synthesis (yellow, light red) and protein decay (black). **d–f**, Cross-sections through **a–c** at time class T5 (broken white lines in **a–c**) for *Kr* (**d**), *Kni* (**e**) and *Gt* (**f**). Broken lines represent relative protein concentration, solid lines represent the rate of change in protein concentration, dv/dt . Domains of protein synthesis and decay are indicated by the light green and red backgrounds, respectively.

effect of *Kni* on *gt* (Fig. 4l). Note that the increasing repression of *kni* by Hb in the posterior portion of the shift zone is a consequence rather than a cause of the *kni* boundary shift (Fig. 4k). It occurs only in nuclei that do not express any *kni* because *Kni* is a very strong repressor of *hb* (Fig. 4c and Supplementary Information).

Figure 3 Gap domain boundary shifts as visualized by asymmetric distribution of transcript (RNA) and protein expression domains. Embryos fluorescently stained for RNA and protein products of *Kr* (**a, g, m**), *kni* (**c, i, o**) and *gt* (**e, k, q**), as well as the

Figure 4 Graphical dynamic analysis of the regulatory mechanism for the shift in the posterior *kni* boundary. Regulatory graphs are shown for *kni* (**a–c, e–g, i–k**) and *gt* (**d, h, l**) over cycles 13 and 14A. **a–c, e–g**, Temporal behaviour of the rate of change in Kni protein production, dW/dt (**a–c**), and the diffusion, and the synthesis and decay terms of the gene circuit equation (**e–g**; see equation (1) in Box 1). Mitosis is shown as a shaded background. The zone of the shift in the posterior *kni* boundary is limited by an anterior nucleus (at 62% A–P position) that shows no Kni protein decay (**a**), and a posterior nucleus (at 80% A–P position) that shows no significant Kni protein synthesis at any time during cycle 14A (**c**). Nuclei in the zone (at 68% A–P position) show a switch from protein synthesis to decay during cycle 14A (arrow, **b, f**). Diffusion counteracts the boundary shift by generating an influx of protein into the zone of protein decay (asterisk, **f**). **d**, Gt

synthesis in the nucleus at 68% A–P position is initiated at early cycle 14A and maintained at low levels throughout cycle 14A (arrowheads). **h**, Influx by diffusion contributes very slightly to the accumulation of Gt protein (plus). **i–l**, Temporal behaviour of regulatory contributions to *kni* (**i–k**) and *gt* (**l**) expression. Vertical axes show relative regulatory contribution. Positive values above the upper broken line represent protein synthesis at more than 90% of the maximum rate, negative values below the lower broken line represent repression, where protein synthesis is less than 10% of the maximum rate (Box 1). The sum of regulatory contributions, u_i is represented by a black line. Coloured areas represent individual regulatory contributions by Hb, Kr, Gt, Kni, Tll, Bcd and Cad to *kni* (**i–k**) and *gt* (**l**). The height of each coloured area is given by $m^a v_i^{\text{Bcd}}$ for Bcd and by $Tab v_b^b$ for all other regulators b (see equation (2) in Box 1).

increasing Kni synthesis in the anterior portion of the shift zone is largely caused by stably maintained Kni autoactivation (Fig. 4i). More posterior nuclei do not reach the threshold concentration of Kni required for stable maintenance of autoactivation and show only moderate transient (Fig. 4j) or no (Fig. 4k) Kni autoactivation. Localized autoactivation is a consequence of the asymmetric repressive mechanism described above, rather than a cause of shifting domain boundaries. This is corroborated by the fact that a gap gene circuit without *kni* autoactivation shows correct shifts of *kni* domain boundaries (see Supplementary Information).

We have also analysed shifts of other gap gene domain boundaries (see Supplementary Information). With the exception of the posterior boundaries of the anterior *hb* and *gt* domains, we have detected anterior shifts in all boundaries examined (Fig. 1c–h). The shifts of the posterior boundaries of *Kr* and posterior *gt* are caused by asymmetric repressive interactions between *Kr* and *kni*, and *gt* and *hb*, very similar to those described for *kni*. Posterior dominance of repressive interactions among neighbouring gap genes is reminiscent of homeotic gene regulation, where posterior homeotic genes repress more anterior ones, but not vice versa¹⁸. Shifts of anterior gap domain boundaries either follow the posterior boundary shifts of more anterior gap genes or are due to sharpening of the posterior boundaries of anterior *gt* and *hb*. For example, the upregulation of *gt* in nuclei anterior to its posterior expression domain follows diminished repression by *Kr* (Fig. 4l). Therefore, shifts of anterior gap domain boundaries can be considered to be secondary effects of the dynamic behaviour of posterior boundaries.

Our results indicate that maternal Bcd, Hb and Cad alone are not sufficient for positioning of gap gene domains and hence do not qualify as morphogens in a strict sense. As has been pointed out¹⁹, an active role of target tissue in specifying positional information contradicts the traditional distinction between the instructive role of maternal morphogens and their passive interpretation¹. The requirement of specific regulatory interactions in the target tissue for proper interpretation of positional information can be interpreted as a requirement for specific tissue competence²⁰. In addition, the dynamical nature of positional information, as encoded by expression boundaries, suggests that positional information in the blastoderm embryo can no longer be seen as a static coordinate system imposed on the embryo by maternal morphogens¹. Rather, it needs to be understood as the dynamic process underlying the positioning of expression domain boundaries, which is based on both external inputs by morphogens and tissue-internal feedback among target genes. □

Methods

Acquisition of quantitative data

Drosophila blastoderm embryos were immunostained for three segmentation gene products each²¹. Each embryo was stained for Even-skipped (Eve) protein for time classification and registration²². We scanned laterally oriented embryos by laser confocal microscopy. Scanned images were aligned along the anteroposterior (A–P) axis and segmented by using binary nuclear masks to yield per-nucleus expression data for each protein.

Processing of quantitative data

Embryos were classified temporally as belonging to cycles 12 (for initial conditions), 13, or eight equally distributed time classes (T1–T8) in cycle 14A. Temporal classification in cycle 14A is based on the highly dynamic expression pattern of *eve*²². Expression patterns were registered by using a fast dyadic wavelet transform²². We removed nonspecific background staining. Data from the middle 10% of dorsoventral (D–V) positional values of each embryo were averaged for each gene and time class to yield an integrated data set that currently contains data based on 954 embryos, available in the FlyEx Database (<http://urchin.spbcas.ru/flyex>).

RNA/protein double staining

RNA was detected using digoxigenin-labelled RNA probes²³. RNA hybridization was done before protein detection by using standard protocols with the exception of permeabilization, where Proteinase K was replaced by acetone treatment²⁴. We obtained quantified expression profiles by image segmentation as described above.

Numerical simulations

Simulator and optimizer code was implemented in C, Java and Perl and is available on the Reinitz Lab website (<http://flyex.ams.sunysb.edu/lab/gaps.html>). Ordinary differential equations were solved numerically by using a Bulirsch–Stoer adaptive step size method²⁵. Integration was done to 0.1% accuracy, and the stability of solutions was confirmed.

Optimization

The sum of squared differences between model and data was minimized using parallel Lam simulated annealing (PLSA)^{10,26}. Search spaces were defined by explicit limits as well as a penalty function¹⁰. Each optimization run was done in parallel on ten 2.4-GHz Pentium P4 Xeon processors and took between 8 and 160 h per gene circuit.

Gene circuit selection and analysis

Because PLSA is a stochastic method, the quality of the resulting gene circuits can vary. Best solutions were selected as described elsewhere¹⁶. The selection process yielded 10 circuits (out of 40), which were used in the present analysis (see Supplementary Information). These circuits show strong constraints toward a specific gap gene network topology¹⁶. All data shown here are based on one particular gap gene circuit that shows no visible patterning defects and corresponds exactly to the network topology observed in most selected circuits.

Received 12 April; accepted 20 May 2004; doi:10.1038/nature02678.

1. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).
2. Crick, F. Diffusion in embryogenesis. *Nature* **225**, 420–422 (1970).
3. Lewis, J., Slack, J. M. W. & Wolpert, L. Thresholds in development. *J. Theor. Biol.* **65**, 579–590 (1977).
4. Gurdon, J. B. & Bourillot, P. Y. Morphogen gradient interpretation. *Nature* **413**, 797–803 (2001).
5. Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
6. Akam, M. The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1–22 (1987).
7. Ingham, P. W. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature* **335**, 25–34 (1988).
8. Mjølness, E., Sharp, D. H. & Reinitz, J. A connectionist model of development. *J. Theor. Biol.* **152**, 429–453 (1991).
9. Reinitz, J., Mjølness, E. & Sharp, D. H. Cooperative control of positional information in *Drosophila* by *bicoid* and maternal *hunchback*. *J. Exp. Zool.* **271**, 47–56 (1995).
10. Reinitz, J. & Sharp, D. H. Mechanism of *eve* stripe formation. *Mech. Dev.* **49**, 133–158 (1995).
11. Reinitz, J., Kosman, D., Vanario-Alonso, C. E. & Sharp, D. H. Stripe forming architecture of the gap gene system. *Dev. Gen.* **23**, 11–27 (1998).
12. Driever, W. & Nüsslein-Volhard, C. The Bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* **54**, 95–104 (1988).
13. Ephrussi, A. & St Johnston, D. Seeing is believing: the Bicoid morphogen gradient matures. *Cell* **116**, 143–152 (2004).
14. Simpson-Brose, M., Treisman, J. & Desplan, C. Synergy between the Hunchback and Bicoid morphogens is required for anterior patterning in *Drosophila*. *Cell* **78**, 855–865 (1994).
15. Houchmandzadeh, B., Wieschaus, E. & Leibler, S. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**, 798–802 (2002).
16. Jaeger, J. *et al.* Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics* (in the press).
17. Gaul, U., Seifert, E., Schuh, R. & Jäckle, H. Analysis of Krüppel protein distribution during early *Drosophila* development reveals posttranscriptional regulation. *Cell* **50**, 639–647 (1987).
18. McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
19. Meinhardt, H. Space-dependent cell determination under the control of a morphogen gradient. *J. Theor. Biol.* **74**, 307–321 (1978).
20. Waddington, C. H. *Organisers and Genes* (Cambridge Univ. Press, Cambridge, UK, 1940).
21. Kosman, D., Small, S. & Reinitz, J. Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Dev. Genes Evol.* **208**, 290–294 (1998).
22. Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M. & Reinitz, J. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics* **17**, 3–12 (2001).
23. Tsai, C. & Gergen, J. P. Gap gene properties of the pair-rule gene *run* during *Drosophila* segmentation. *Development* **120**, 1671–1683 (1994).
24. Nagaso, H., Murata, T., Day, N. & Yokoyama, K. K. Simultaneous detection of RNA and protein by *in situ* hybridization and immunological staining. *J. Histochem. Cytochem.* **49**, 1177–1182 (2001).
25. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C* (Cambridge Univ. Press, Cambridge, UK, 1992).
26. Chu, K. W., Deng, Y. & Reinitz, J. Parallel simulated annealing by mixing of states. *J. Comp. Phys.* **148**, 646–662 (1999).
27. Foe, V. E. & Alberts, B. M. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila*. *J. Cell Sci.* **61**, 31–70 (1983).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. P. Gergen for the constructs for RNA probes; and N. Monk, J. Dallman, J. D. Baker and L. Carey for comments on the manuscript. This work was supported financially by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R. (reinitz@odd.bio.sunysb.edu).

Chromatin regulates origin activity in *Drosophila* follicle cells

Bhagwan D. Aggarwal & Brian R. Calvi

Department of Genetics, University of Pennsylvania School of Medicine, 415 Curie Blvd, Philadelphia, Pennsylvania 19104, USA

It is widely believed that DNA replication in multicellular animals (metazoa) begins at specific origins to which a pre-replicative complex (pre-RC) binds¹. Nevertheless, a consensus sequence for origins has yet to be identified in metazoa. Origin identity can change during development, suggesting that there are epigenetic influences. A notable example of developmental specificity occurs in *Drosophila*, where somatic follicle cells of the ovary transition from genomic replication to exclusive re-replication at origins that control amplification of the eggshell (chorion) protein genes². Here we show that chromatin acetylation is critical for this developmental transition in origin specificity. We find that histones at the active origins are hyperacetylated, coincident with binding of the origin recognition complex (ORC). Mutation of the histone deacetylase (HDAC) *Rpd3* induced genome-wide hyperacetylation, genomic replication and a redistribution of the origin-binding protein ORC2 in amplification-stage cells, independent of effects on transcription. Tethering *Rpd3* or Polycomb proteins to the origin decreased its activity, whereas tethering the Chameau acetyltransferase increased origin activity. These results suggest that nucleosome acetylation and other epigenetic changes are important modulators of origin activity in metazoa.

To explore the relationship between chromatin modification and metazoan origin activity, we examined the origins that control developmental amplification of the *Drosophila* chorion genes. During stage 10A of oogenesis, somatic follicle cells undergo a developmental transition from genomic replication to continuous re-replication from origins at the chorion loci on the X and 3rd chromosome (hereafter referred to as X and 3rd chorion), as well as at two other recently identified loci^{2–5}. Beginning at stage 10B of oogenesis, this continuous re-replication can be seen as four sub-nuclear foci by 5-bromodeoxyuridine (BrdU), fluorescence *in situ* hybridization (FISH), or antibody labelling for replication proteins^{2–6–8}. The regulation of chorion origins resembles normal cell cycle control in that they assemble a pre-RC, which is activated by S-phase kinases⁹. The 3rd chorion locus amplifies to highest copy number, and the two sequences that are most important for this are termed amplification controlling element 3 (ACE3) and Ori- β , which bind to ORC *in vitro* and *in vivo*¹⁰ (Fig. 1a). These sequences can direct amplification when inserted at ectopic chromosomal sites, but the level of amplification is extremely sensitive to genomic position¹¹, which can be buffered by chromatin insulators¹². This is consistent with the notion that the chromatin neighbourhood in which the origin resides can have a major influence on its activity.

To evaluate whether chromatin is modified at chorion origins, we immunolabelled *Drosophila* ovaries with antibodies against poly-acetylated histone H4 (AcH4). We detected four prominent nuclear foci in amplification-stage follicle cells, with a lower level of staining throughout the nucleus (Fig. 1b). Several observations suggested that these four hyperacetylation foci correspond to origins at amplification loci during active initiation. First, the four hyperacetylation foci were coincident with amplification foci detected by an antibody to ORC2 (Fig. 1b–d). Second, at the 3rd chorion locus both ORC2 and AcH4 were localized to the origin during the period of active initiations from early stage 10B to late stage 11, but were not seen during the majority of stages 12–14 (a

time when no new initiations occur but forks continue to migrate)^{6,10,13} (Supplementary Fig. S1). However, AcH4 did persist slightly later than ORC2 into early stage 12 (a difference of less than one hour in developmental time). Third, hyperacetylation at the 3rd chorion locus corresponded closely to ORC2 labelling at the origin, but did not overlap with labelling for Double-parked protein (Dup), which co-localizes with migrating replication forks. This further suggested that AcH4 labelling does not represent acetylation on newly deposited nucleosomes behind forks (Fig. 1e–g and Supplementary Fig. S2)¹³. Although amplification increased the prominence of AcH4 foci, their intensity could not be accounted for by DNA copy number alone (Supplementary Figs S2 and S3). Similar results were obtained with antibodies against poly-acetylated histone H3 and histone H4 acetylated on lysine 8 (Supplementary Fig. S4).

To delimit histone modification near the origin with higher resolution, we performed chromatin immunoprecipitation (ChIP) with an AcH4 antibody on genomic DNA from amplification-stage egg chambers. This indicated that ACE3 and Ori- β were both enriched approximately 25-fold in the AcH4 precipitation, relative to a non-amplified control locus, whereas sequences 10–50 kilobases (kb) proximal and distal to the 3rd chorion locus were not enriched (Fig. 2). This suggests that nucleosomes are hyperacetylated at ACE3 and Ori- β , the two ORC binding sites critical for amplification.

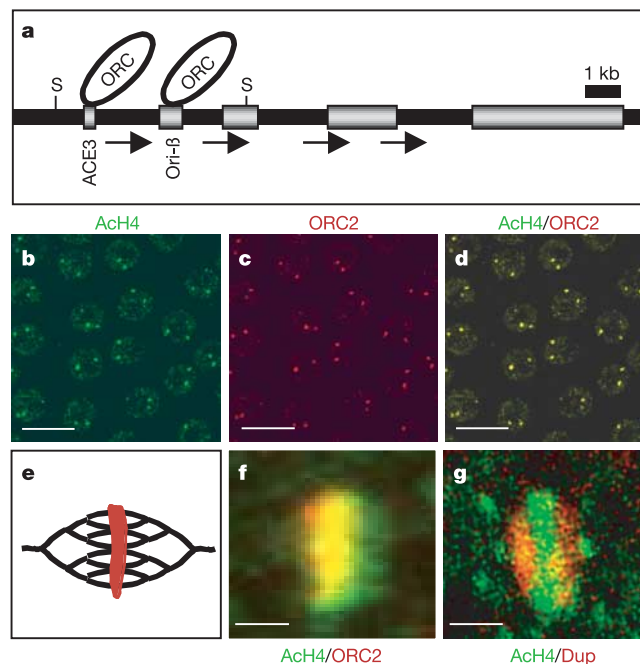


Figure 1 Histone hyperacetylation and ORC2 co-localize at chorion origins. **a**, Organization of the 3rd chromosome chorion locus. Arrows represent the four chorion genes, grey boxes represent the five regions that contribute to amplification. The two most important, ACE3 and Ori- β are binding sites for ORC. Sa/I sites (S) define the 3.8-kb fragment used in subsequent experiments. **b–d**, Labelling of stage-10B follicle cells with poly-acetylated histone H4 (AcH4) antibody (green; **b**), ORC2 (red; **c**), and merge (yellow; **d**). The two brightest foci represent the chorion locus on the 3rd and X chromosome. **e**, Representation of ORC (red) binding to ACE3/Ori- β on amplified chorion DNA fibres based on previous reports^{6,10,13}. **f**, High magnification showing co-localization of ORC2 (red) and AcH4 (green) at the 3rd chorion origin in stage 11. **g**, AcH4 (green) does not co-localize with Dup protein (red) which labels replication forks that migrate bi-directionally outward from the 3rd chorion origin¹³. Scale bars represent 10 μ m (**b–d**) and 3 μ m (**f, g**).

To determine whether acetylation regulates origin activity, we examined loss-of-function mutations in the HDAC gene *Rpd3*. In yeast, *Rpd3* mutants advance the time at which late origins fire, a relatively subtle effect compared with the effect on transcription¹⁴. Severe mutations in *Drosophila Rpd3* are homozygous lethal¹⁵. Therefore, we used the FLP/FRT recombination system to create clones of follicle cells homozygous for the strong *Rpd3^{m5-5}* allele (J. Simon and M. O' Connor, personal communication). To avoid pleiotropic effects that result from prolonged depletion of *Rpd3* protein, we generated small clones (1–10 cells) late in oogenesis. Labelling with a *Rpd3* antibody confirmed that cells within these clones had reduced levels of *Rpd3* protein (Supplementary Fig. S5).

Consistent with a defect in an HDAC, labelling with AcH4 and other antibodies indicated that amplification-stage *Rpd3*-mutant follicle cells had a 3 to 4 fold increase in hyperacetylation throughout the nucleus (Fig. 3a, b and Supplementary Fig. S6). Approximately 85% ($n = 100$) of small clones with global acetylation also had altered replication, with BrdU incorporation throughout the nucleus instead of just at the amplification loci as seen in neighbouring wild-type cells (Fig. 3c, d and Supplementary Fig. S7). To test whether BrdU labelling in *Rpd3*-mutant cells could represent a change in origin usage, we examined the distribution of ORC2. Instead of the normal focal staining at chorion loci, many *Rpd3*-mutant cells with hyperacetylation also had a low level of

punctate ORC2 labelling throughout the nucleus, with smaller clones most frequently having redistributed ORC2 (Fig. 3e–j and Supplementary Fig. S8). Among *Rpd3*-mutant clones comprised of five or fewer cells, 20% ($n = 41$) had at least one large nucleus, and measurement of total 4,6-diamidino-2-phenylindole (DAPI) fluorescence indicated that they contained approximately twofold more DNA than neighbouring non-mutant cells (Fig. 3f, h, j, asterisk and data not shown). This suggests that they had undergone an extra genome reduplication rather than a developmental delay in genomic replication that normally occurs before stage 10B. In most cells that incorporated BrdU, however, we could not detect a significant increase in DNA content, suggesting that ongoing replication had not resulted in a full reduplication of the genome. Similar cell–cell variation has been noted in a number of other mutants that alter replication in follicle cells^{16,17}.

Treatment of egg chambers *in vitro* with the HDAC inhibitors sodium butyrate or trichostatin A (TSA) also resulted in hyperacetylation and extra genomic BrdU incorporation in stage-10B follicle cells (Fig. 4a, b, g and data not shown). This included a notable increase in BrdU incorporation within the heterochromatic chromocentre, which is under-replicated in most endocycles¹⁸ (Fig. 4b). The increased BrdU incorporation into this nuclear compartment, which is known to differ in histone acetylation, suggests that HDAC inhibitors, and mutation of *Rpd3*, alter replication specificity by changing acetylation status.

In the cells with hyperacetylation, it is possible that the altered replication specificity is due to an increase in replication-protein expression, rather than a direct effect on chromatin at the origin¹⁹. Measurement of the total fluorescence intensity of redistributed ORC2 in *Rpd3*-mutant cells, however, indicated that it was not significantly increased (Supplementary Fig. S8 and data not shown). To answer this question, we asked if the transcription inhibitor α -amanitin would block the extra replication that results from inhibition of HDACs with sodium butyrate¹⁶. Although α -amanitin pre-treatment strongly inhibited heat induction of an hsp70:Myc-tagged reporter in controls, it did not have a significant effect on the extra genomic replication seen in sodium-butyrate-treated egg chambers (Fig. 4a–g). These results suggest that the extra genomic replication after treatment with HDAC inhibitors and in *Rpd3*-mutant cells is not mediated solely by increased transcription of replication-protein genes.

To test further whether chromatin modification at the origin can have an impact on replication activity locally, we established a two-part system in which chromatin-modifying proteins are tethered to the 3rd chorion origin *in vivo* (Fig. 5a). We transformed flies with an amplification reporter, Tether Target 1 (TT1), which contains five copies of GAL4-binding sites adjacent to the minimal 3rd chorion origin (which contains the ORC binding sites ACE3 and Ori- β ; Figs 1a and 5a). Other strains were transformed with *Rpd3* fused to the GAL4 DNA binding domain (GAL4^{DBD}), and expressed under control of the heat-inducible hsp70 promoter (hsp70:GAL4^{DBD}:*Rpd3*) (Fig. 5a and data not shown). In females containing both transgenes, we measured the ratio of copy number for TT1 versus the endogenous 3rd chorion locus within each follicle cell nucleus by FISH, using an origin probe².

Heat induction of hsp70:GAL4^{DBD}:*Rpd3* expression reduced TT1 amplification to 59% of controls ($P < 0.0001$; Fig. 5d, e, j). Expression of GAL4^{DBD} alone, or heat-treatment alone, had no specific effect on TT1 amplification ($P \approx 0.6$; Fig. 5b, c, j). Moreover, hsp70:GAL4^{DBD}:*Rpd3* expression had no effect on S6.9-5, a 3rd chorion P element that lacks GAL4-binding sites (data not shown). In some cells, expression of hsp70:GAL4^{DBD}:*Rpd3* reduced TT1 amplification to below the level of detection by FISH. Therefore, we also used the fraction of nuclei with detectable TT1 and endogenous 3rd chorion signal as a measure of relative origin strength. This gave similar results, with hsp70:GAL4^{DBD}:*Rpd3* reducing detection to 70% of controls (Fig. 5j). Finally, quantitative

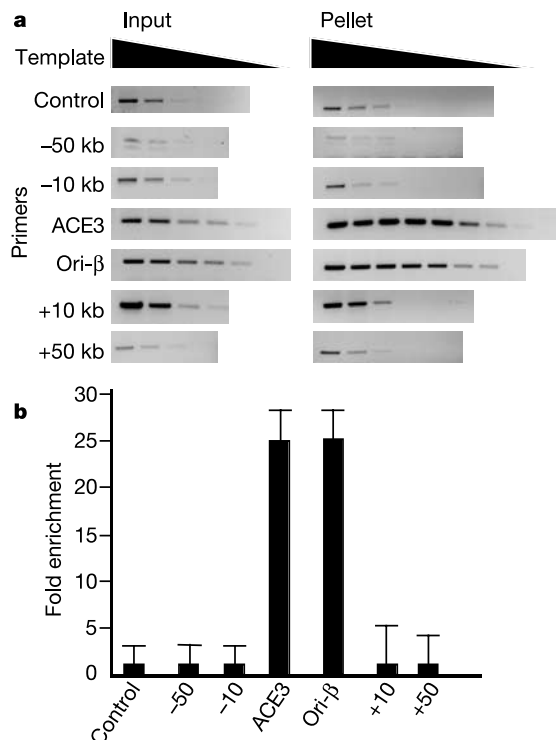


Figure 2 ChIP analysis indicates that nucleosomes are hyperacetylated at the 3rd chromosome chorion origin. ChIP from stage-10B and stage-11 egg chambers using anti-polyacetylated H4 (AcH4). **a**, Fivefold dilutions of input or pellet template DNA were subjected to PCR using primers to 3rd chorion ORC binding sites (ACE-3 and Ori- β ; see Fig. 1a), sequences 10 kb and 50 kb proximal and distal to the origin, and a non-amplified control at 6C (control). PCR products were separated on agarose gels and stained with ethidium bromide. **b**, Quantification of PCR products. The ratio of pellet to input was normalized to the pellet to input ratio for the control. The bar graph represents the average and standard deviations of normalized enrichment for two independent precipitations and multiple PCR reactions.

Southern blotting of stage-13 egg chambers suggested that hsp70–GAL4^{DBD}–Rpd3 reduces TT1 amplification to 75% of controls (Fig. 5j). Similar results were obtained with another TT1 inserted at a different genomic location (data not shown). This suggests that Rpd3 can inhibit origin activity locally.

To evaluate the contribution of HDAC activity to repression of the origin, we tethered an Rpd3 protein mutated in an active-site histidine that is essential for HDAC activity in other organisms²⁰. This mutant, GAL4^{DBD}:Rpd3^{H137A}, retained partial repression activity and reduced TT1 amplification to 66% ($P < 0.0001$), and reduced detection to 89%. This is consistent with results from yeast and mammalian cells, where catalytically inactive GAL4^{DBD}:Rpd3 fusions can partially repress transcription *in vivo*²⁰. Similarly to the results at promoters, it is likely that Rpd3^{H137A} can recruit other repressing proteins to the origin and, therefore, does not distinguish the contribution of deacetylation to origin repression.

Therefore, we tethered HAT1, a product of the *chameau* gene and the putative fly orthologue of human HBO1, to TT1^{21,22}. HBO1 has been shown to associate with the *Drosophila* and human pre-RC proteins MCM2 and ORC1 (refs 21, 23). Expression of hsp70:GAL4^{DBD}:HAT1 increased the ratio of TT1/3rd to 153% ($P < 0.0001$), increased the fraction of nuclei with detectable TT1 to 126% and increased the copy number on Southern blots to 165% of controls (Fig. 5f, g, j). Similar results were obtained with a TT1 inserted at a different genomic location (data not shown). This suggests that increased acetylation stimulates origin activity.

We also asked whether a repressive chromatin environment mediated by the silencing protein Polycomb (Pc) diminishes origin activity. Polycomb is part of a multi-protein complex that associates with HDACs, and is required for maintenance of chromatin domains that repress transcription at homeotic and other loci²⁴. Tethering Pc to TT1 reduced the TT1/3rd ratio to 66%

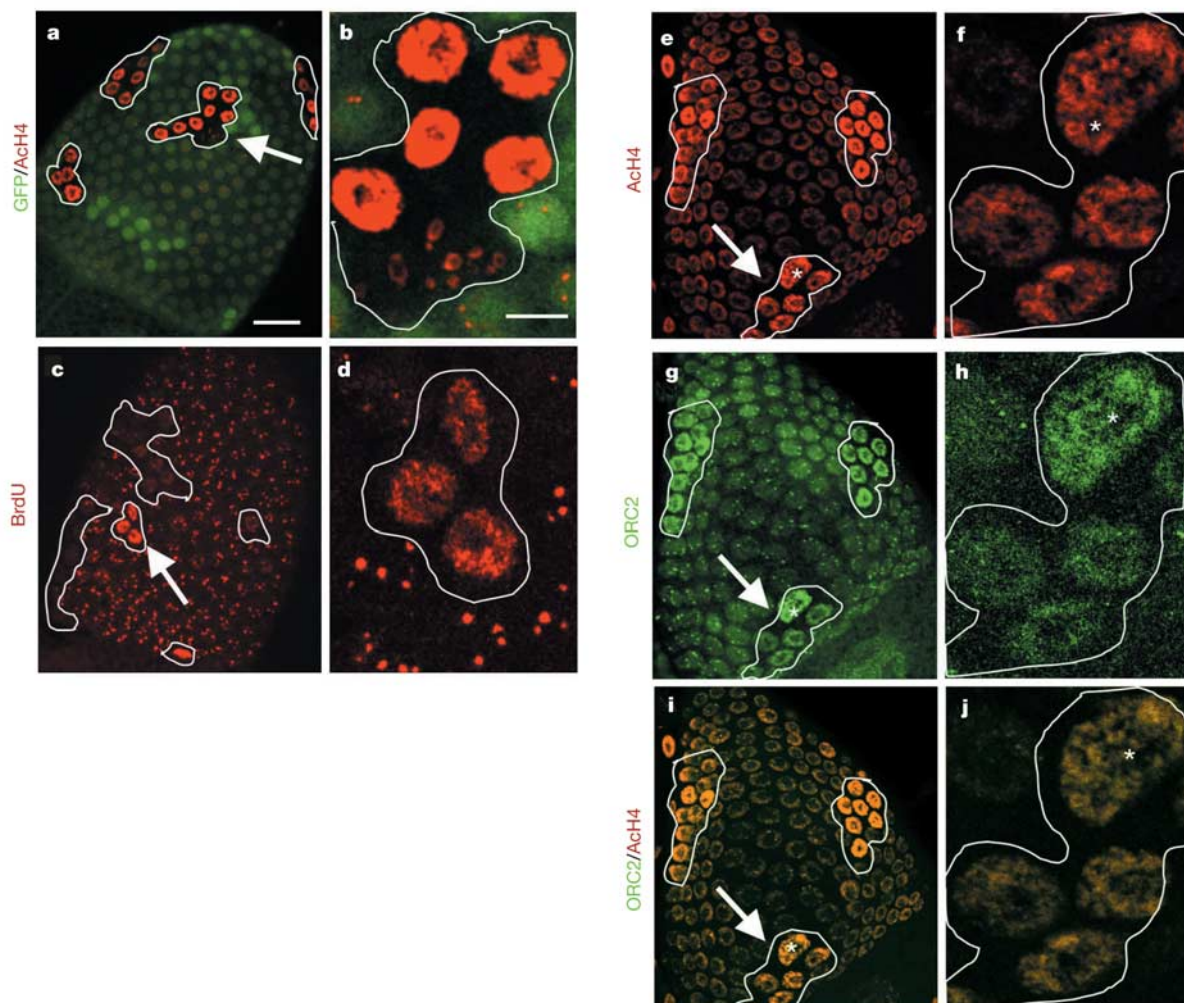


Figure 3 *Rpd3* loss-of-function clones show hyperacetylation, increased replication and altered ORC2 distribution. Homozygous mutant *Rpd3*^{m5-5}-follicle-cell clones were generated using the FLP/FRT technique. **a, c, e, g, i**, Low magnification of stage-10B egg chambers. **b, d, f, h, j**, Higher magnification images of clones designated by arrows. **a, b**, Mutant clones were identified in stage-10B egg chambers by the absence of green fluorescent protein (GFP) fluorescence (green), and labelling for Ach4 indicated that most *Rpd3*^{m5-5} mutant cells within small clones have nucleus-wide hyperacetylation (red). Rare regional hyperacetylation can be seen for one cell in the lower part of the clone in (**b**). **c, d**, Many cells in small clones also had inappropriate nucleus-wide incorporation of

BrdU, instead of the focal staining at chorion loci seen in neighbouring wild-type cells (see Supplementary Fig. S7). **e–j**, Orc2, Ach4 double labelling. Acetylation (red; **e, f**), ORC2 (green; **g, h**), merge (yellow; **i, j**). ORC2 was distributed throughout the nucleus in 51% of clones comprised of ten or fewer cells ($n = 242$ clones; see Supplementary Fig. S8). Note that outside the clones the green colour is from ORC2 and GFP. The asterisk in **f, h** and **j** indicates a nucleus that contains twice the DNA content of its neighbours, as measured by total DAPI fluorescence. Scale bars represent 20 μm (**a, c, e, g, i**) and 10 μm (**b, d, f, h, j**).

($P < 0.0001$), reduced the fraction of nuclei with two FISH signals to 65% and reduced the copy number on Southern blots to 49% of controls (Fig. 5h, i, j). These results suggest that Rpd3 and Pc can inhibit origin activity locally, whereas HAT1 can stimulate it.

Our results suggest that epigenetic modification of chromatin, mediated in part by *Rpd3*, contributes to differential origin usage during the developmental transition from genomic replication to amplification in follicle cells. These results are consistent with results in yeast, where mutation of *Rpd3* advances the time of late origin firing, and previous reports that pre-RC proteins associate with HATs in a number of organisms^{14,21,23,25}. Although chromatin modification may have an important influence, we suggest that it is not the only property that determines origin identity, and that binding of other proteins to specific sites also contributes. In the *Rpd3*-mutant cells, widespread acetylation may open chromatin and permit origin proteins to gain access to additional binding sites, thereby imparting origin activity. It remains unclear, however, at what step(s) acetylation influences origin activity. Acetylation at the chorion locus is coincident with the period when ORC must bind to newly amplified fibres, and genomic hyperacetylation resulted in a redistribution of ORC2, suggesting effects on ORC recognition of DNA. However, acetylation of chromatin, and perhaps of the pre-RC proteins themselves, may also stimulate pre-RC assembly and activation downstream of ORC binding. A rigorous test of this

model awaits further characterization of additional origins in *Drosophila*.

Like *Rpd3*, the *Drosophila* retinoblastoma-family protein, *Rbf1*, represses the activity of chorion and genomic origins in follicle

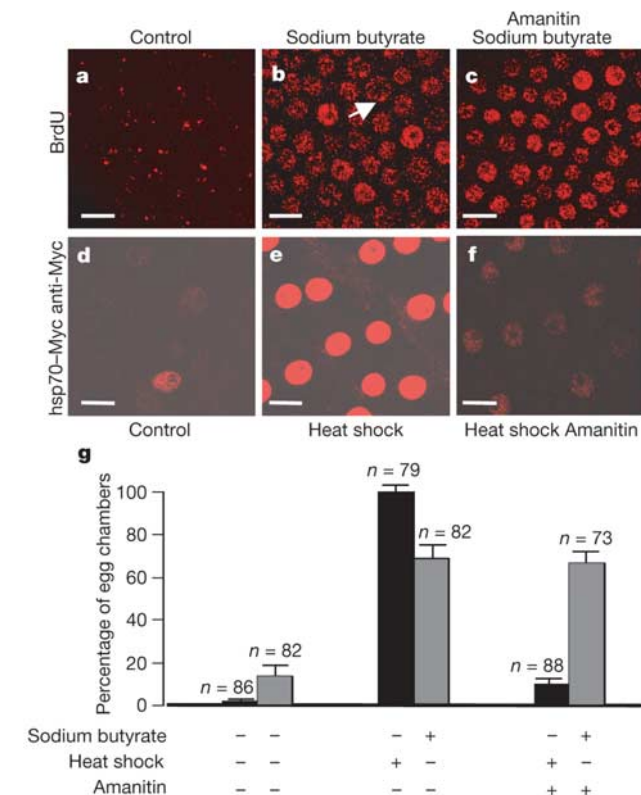


Figure 4 Sodium butyrate induces extra replication, which is not blocked by α -amanitin. **a–c**, BrdU incorporation in stage-10B follicle cells (untreated, **a**; sodium butyrate, **b**; α -amanitin and sodium butyrate, **c**). The arrow in **b** points to intense focal incorporation in one chromocentre. **d–f**, Anti-Myc labelling as a control for the effectiveness of α -amanitin to inhibit transcription of a strong *hsp70*:Myc reporter in follicle cells (no heat induction, **d**; heat induction without α -amanitin, **e**; heat induction with α -amanitin, **f**). **g**, The percentage of total egg chambers that had genomic BrdU incorporation (grey bars) or Myc labelling (black bars) after the indicated treatments. Bars represent average and standard deviation of three experiments; scale bars represent 10 μ m.

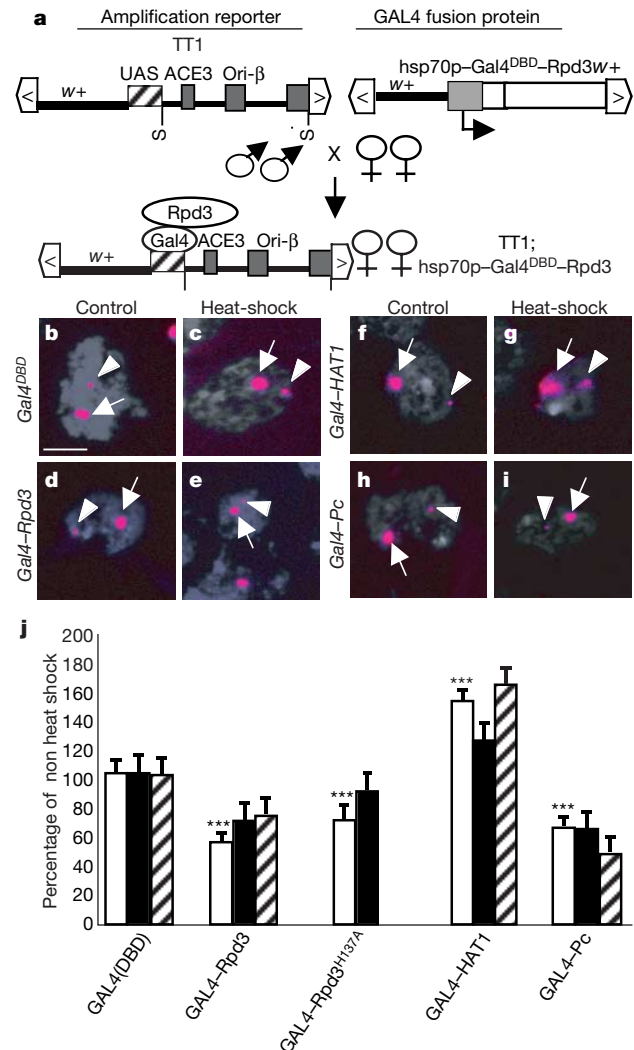


Figure 5 Tethering chromatin modifiers to chorion origins alters their activity. **a**, Method to tether chromatin modifiers to the chorion origin. TT1 contains the 3.8-kb minimal origin from the 3rd chromosome chorion locus, which contains ACE3 and Ori- β (dark-grey boxes; see Fig. 1a), adjacent to GAL4-binding sites (UAS, striped box), and is marked with the *mini-white* gene (*w+*). GAL4 fusions contain the heat-inducible *hsp70* promoter (light-grey box) driving expression of the GAL4-DNA-binding domain (DBD) to Rpd3 (white boxes), or other chromatin modifiers. Boxes with arrowheads are the inverted repeats of the P element transformation vector. Heat-induced expression in progeny with both P elements results in binding of GAL4:Rpd3 to the UAS sites next to the origin in TT1. **b–i**, Single nuclei with representative FISH labelling of the endogenous 3rd chromosome chorion locus (arrow) and TT1 (arrowhead). **b, c**, GAL4^{DBD} only. **d, e**, Rpd3 fusion. **f, g**, HAT1 fusion. **h, i**, Polycomb fusion. Uninduced (**b, d, f, h**), induced (**c, e, g, i**) GAL4 fusion. The variable nuclear morphology results from the squashing technique for FISH. **j**, Percentage of amplification for TT1 relative to uninduced sibling controls that were normalized to 100%. Females contained GAL4^{DBD} alone or the indicated fusions. TT1 levels were measured by: the average ratio of TT1 to endogenous fluorescence intensity (white bars); the fraction of nuclei with endogenous 3rd chorion labelling that had a detectable TT1 spot (black bars); and Southern blotting (striped bars). Each bar represents the average and standard error of the mean for three to four independent experiments. *** $P < 0.0001$ for mean of TT1/3rd locus ratio between induced and uninduced controls by *t* test. Scale bar represents 5 μ m (**b–i**).

cells¹⁶. Mammalian HDACs that are similar to Rpd3 associate with Rb to mediate transcriptional repression²⁶, and Rb associates with replication foci and origins in mammalian cells^{27,28}. It is possible, therefore, that during amplification Rpd3 mediates a repressive effect of Rb at most genomic origins, which may be counteracted by the chameau HAT at chorion origins. It was recently shown that another protein known to associate with chromatin modifiers, the *Drosophila* Myb protein, binds ACE3 and is required for normal amplification²⁹. The ability of transcriptional regulators to recruit enzymes that modify and remodel chromatin, therefore, may be one of the critical determinants for where and when replication starts on metazoan chromosomes. □

Methods

Plasmid construction

Details of plasmid construction are described in the Supplementary Information.

Fly strains

Information on genetic nomenclature and strains can be found at <http://flybase.bio.indiana.edu>. The FRT strain $w^+P\{w^{+}, hsp70-FLP\}; Rpd3^{MS-5}P\{w^{+}, m^{w}hs=FRT(w^{hs})\}2A$ was a gift from J. Simon and M. O'Connor. Strain $P\{w^{+}, m^{w}hs=FRT(w^{hs})\}2A$ was a gift from D. Beuchle. For most experiments, flies were raised at 22 °C on standard cornmeal media.

Tethering assay and FISH

Females were conditioned on wet yeast for three days. Hsp70:GAL4 fusions were induced at 37 °C for 30 min in a water bath. Siblings were left uninduced as controls. The egg chambers were dissected and processed for FISH or genomic Southern blot 6 h after heat induction. Southern blot quantifications and FISH were as described². FISH images were captured with a Leica DMR and Hamamatsu CCD and quantified using OpenLab software (Improvision). The ratio of TT1 signal to endogenous locus was calculated in each nucleus and averaged over at least 40 nuclei for each of three independent experiments. We included in the calculation only nuclei in which both TT1 and endogenous 3rd chorion could be detected. Because we observed different amplification levels due to genetic background, these values were normalized to the TT1/endogenous ratio measured in uninduced sibling controls. Alternatively, the fraction of nuclei with a detectable TT1 hybridization among nuclei with a detectable 3rd chorion signal was calculated. An unpaired *t* test was used to evaluate the significance of the difference between control and experimental samples.

Immunofluorescence microscopy

Methods for fixation and analysis of ovaries were as described². Mouse anti-ORC2 (clone no. 81E7; a gift from M. Botchan), was used at a dilution of 1:20. Antibodies against poly-Ach4, poly-Ach3, AcK5H4, AcK8H4, and AcK12H4 (Upstate) were used at a dilution of 1:200. Polyclonal rabbit anti-Rpd3 was a gift from J. Kadonaga and was used at a dilution of 1:1,000. Secondary antibodies, Cy3 anti-mouse (Jackson Immuno Research) and Alexa-488 anti-rabbit (Molecular Probes) were used at 1:400. Antibody labelling was analysed by Leica DMR confocal and TCS-NT software. BrdU and DAPI labelling were as described².

Chromatin immunoprecipitation

ChIP from stage 10B–11 egg chambers was as described¹⁰. Either 10 µl of rabbit Anti-Ach4 (Upstate) or rabbit pre-immune serum was added to the supernatant for 4 h before addition of 60 µl salmon sperm DNA/protein A agarose beads (Upstate) for an additional 1 h. All post-immunoprecipitation procedures were performed as described in the Upstate Ach4 ChIP assay kit protocol. ChIP polymerase chain reactions (PCRs) contained input or pellet DNA diluted in fivefold increments. PCR products were analysed on ethidium-bromide-stained agarose gels and quantified against a standard curve using a MultiImager (Alpha Innotech Corporation) and ChemilMager V5.5 (Alpha Innotech Corporation) software. Figure 2 represents the averages and standard deviations for two independent immunoprecipitations and four PCR reactions for Ori-β and ACE3, and three to four PCR reactions for other 3rd chromosome primers.

FRT-Rpd3 clones

Homozygous $Rpd3^{MS-5}$ -follicle-cell clones were generated by standard mitotic recombination using the FLP/FRT system. $w^+P\{w^{+}, hsp70-FLP\}; Rpd3^{MS-5}P\{w^{+}, m^{w}hs=FRT(w^{hs})\}2A/P\{w^{+}, m^{w}hs=Ubi-GFP\}61EF P\{w^{+}, m^{w}hs=FRT(w^{hs})\}2A$ females were conditioned on wet yeast for 3 d. To induce FLPase, females were incubated at 38 °C for 40 min and transferred back to wet yeast at 22 °C. Females were dissected 24–72 h after heat shock.

Treatment of egg chambers with sodium butyrate/α-amanitin *in vitro*

Ovaries were pre-incubated in Grace's insect cell culture medium (Cellgro) with or without 200 µg ml⁻¹ α-amanitin (Sigma) for 15 min at 30 °C with rocking. Egg chambers were incubated with or without 10 mM sodium butyrate (Upstate) in Grace's for 1 h, before 1 h in Grace's without sodium butyrate at 30 °C, with or without α-amanitin. Egg chambers were labelled for BrdU as described, with or without α-amanitin until the egg chambers were fixed. The control hsp70:Myc reporter egg chambers were pre-incubated

with or without amanitin for 15 min as above, and heat treated at 37 °C for 1 h, before a 2-h recovery at room temperature and fixation and immunolabelling for Myc.

Received 8 March; accepted 27 May 2004; doi:10.1038/nature02694.

- Bell, S. P. & Dutta, A. DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* **71**, 333–374 (2002).
- Calvi, B. R., Lilly, M. A. & Spradling, A. C. Cell cycle control of chorion gene amplification. *Genes Dev.* **12**, 734–744 (1998).
- Claycomb, J. M., Benasutti, M., Bosco, G., Fenger, D. D. & Orr-Weaver, T. L. Gene amplification as a developmental strategy: isolation of two developmental amplicons in *Drosophila*. *Dev. Cell* **6**, 145–155 (2004).
- Spradling, A. & Mahowald, A. Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **77**, 1096–1100 (1980).
- Calvi, B. R. & Spradling, A. C. Chorion gene amplification in *Drosophila*: A model for metazoan origins of DNA replication and S-phase control. *Methods* **18**, 407–417 (1999).
- Calvi, B. R. & Spradling, A. C. The nuclear location and chromatin organization of active chorion amplification origins. *Chromosoma* **110**, 159–172 (2001).
- Royzman, I., Austin, R. J., Bosco, G., Bell, S. P. & Orr-Weaver, T. L. ORC localization in *Drosophila* follicle cells and the effects of mutations in *dE2F* and *dDP*. *Genes Dev.* **13**, 827–840 (1999).
- Whittaker, A. J., Royzman, I. & Orr-Weaver, T. L. *Drosophila* double parked: a conserved, essential replication protein that colocalizes with the origin recognition complex and links DNA replication with mitosis and the down-regulation of S phase transcripts. *Genes Dev.* **14**, 1765–1776 (2000).
- Bandura, J. L. & Calvi, B. R. Duplication of the genome in normal and cancer cell cycles. *Cancer Biol. and Ther.* **1**, 8–13 (2002).
- Austin, R. J., Orr-Weaver, T. L. & Bell, S. P. *Drosophila* ORC specifically binds to ACE3, an origin of DNA replication control element. *Genes Dev.* **13**, 2639–2649 (1999).
- de Cicco, D. & Spradling, A. Localization of a *cis*-acting element responsible for the developmentally regulated amplification of *Drosophila* chorion genes. *Cell* **38**, 45–54 (1984).
- Lu, L., Zhang, H. & Tower, J. Functionally distinct, sequence-specific replicator and origin elements are required for *Drosophila* chorion gene amplification. *Genes Dev.* **15**, 134–146 (2001).
- Claycomb, J. M., MacAlpine, D. M., Evans, J. G., Bell, S. P. & Orr-Weaver, T. L. Visualization of replication initiation and elongation in *Drosophila*. *J. Cell Biol.* **159**, 225–236 (2002).
- Vogelauer, M., Rubbi, L., Lucas, L., Brewer, B. J. & Grunstein, M. Histone acetylation regulates the time of replication origin firing. *Mol. Cell* **10**, 1223–1233 (2002).
- Mottus, R., Sobel, R. E. & Grigliatti, T. A. Mutational analysis of a histone deacetylase in *Drosophila melanogaster*: missense mutations suppress gene silencing associated with position effect variegation. *Genetics* **154**, 657–668 (2000).
- Bosco, G., Du, W. & Orr-Weaver, T. L. DNA replication control through interaction of E2F-RB and the origin recognition complex. *Nature Cell Biol.* **3**, 289–295 (2001).
- Schwed, G., May, N., Pechersky, Y. & Calvi, B. R. *Drosophila* minichromosome maintenance 6 is required for chorion gene amplification and genomic replication. *Mol. Biol. Cell* **13**, 607–620 (2002).
- Lilly, M. & Spradling, A. The *Drosophila* endocycle is controlled by cyclin E and lacks a checkpoint ensuring S-phase completion. *Genes Dev.* **10**, 2514–2526 (1996).
- Cayirlioglu, P., Ward, W. O., Silver Key, S. C. & Duronio, R. J. Transcriptional repressor functions of *Drosophila* E2F1 and E2F2 cooperate to inhibit genomic DNA synthesis in ovarian follicle cells. *Mol. Cell Biol.* **23**, 2123–2134 (2003).
- Kadosh, D. & Struhl, K. Histone deacetylase activity of Rpd3 is important for transcriptional repression *in vivo*. *Genes Dev.* **12**, 797–805 (1998).
- Iizuka, M. & Stillman, B. Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. *J. Biol. Chem.* **274**, 23027–23034 (1999).
- Grienerberger, A. *et al.* The MYST domain acetyltransferase Chameau functions in epigenetic mechanisms of transcriptional repression. *Curr. Biol.* **12**, 762–766 (2002).
- Burke, T. W., Cook, J. G., Asano, M. & Nevins, J. R. Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. *J. Biol. Chem.* **276**, 15397–15408 (2001).
- Simon, J. A. & Tamkun, J. W. Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr. Opin. Genet. Dev.* **12**, 210–218 (2002).
- Takei, Y. *et al.* MCM3AP, a novel acetyltransferase that acetylates replication protein MCM3. *EMBO Rep.* **2**, 119–123 (2001).
- Luo, R. X., Postigo, A. A. & Dean, D. C. Rb interacts with histone deacetylase to repress transcription. *Cell* **92**, 463–473 (1998).
- Kennedy, B. K., Barbie, D. A., Classon, M., Dyson, N. & Harlow, E. Nuclear organization of DNA replication in primary mammalian cells. *Genes Dev.* **14**, 2855–2868 (2000).
- Avni, D. *et al.* Active localization of the retinoblastoma protein in chromatin and its response to S phase DNA damage. *Mol. Cell* **12**, 735–746 (2003).
- Beall, E. L. *et al.* Role for a *Drosophila* Myb-containing protein complex in site-specific DNA replication. *Nature* **420**, 833–837 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Simon and M. O'Connor for FRT Rpd3 flies; D. Buechle for hsp70:Gal4:PC flies; M. Botchan, L. Beall, T. Orr-Weaver and S. Bell for ORC2 and Dup antibodies; D. Fyodorov and J. Kadonaga for Rpd3 antibody. We thank J. Claycomb for advice about QPCR. Thanks to N. May and M. Thomer for help with injections and flies. We are indebted to J. Bandura, A. K. Bielsky, and M. Lilly for comments on the manuscript. This work was supported by a PHS grant to B.R.C.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.R.C. (Calvi@mail.med.upenn.edu).

Replication by human DNA polymerase- ϵ occurs by Hoogsteen base-pairing

Deepak T. Nair¹, Robert E. Johnson², Satya Prakash², Louise Prakash² & Anel K. Aggarwal¹

¹Structural Biology Program, Department of Physiology and Biophysics, Mount Sinai School of Medicine, Box 1677, 1425 Madison Avenue, New York, New York 10029, USA

²Sealy Center for Molecular Science, University of Texas Medical Branch, 6.014 Medical Research Building, 11th and Mechanic Streets, Galveston, Texas 77755, USA

Almost all DNA polymerases show a strong preference for incorporating the nucleotide that forms the correct Watson–Crick base pair with the template base. In addition, the catalytic efficiencies with which any given polymerase forms the four possible correct base pairs are roughly the same. Human DNA polymerase- ϵ (hPol ϵ), a member of the Y family of DNA polymerases, is an exception to these rules. hPol ϵ incorporates the correct nucleotide opposite a template adenine with a several hundred to several thousand fold greater efficiency than it incorporates the correct nucleotide opposite a template thymine, whereas its efficiency for correct nucleotide incorporation opposite a template guanine or cytosine is intermediate between these two extremes^{1–5}. Here we present the crystal structure of hPol ϵ bound to a template primer and an incoming nucleotide. The structure reveals a polymerase that is ‘specialized’ for Hoogsteen base-pairing, whereby the templating base is driven to the *syn* conformation. Hoogsteen base-pairing offers a basis for the varied efficiencies and fidelities of hPol ϵ opposite different template bases, and it provides an elegant mechanism for promoting replication through minor-groove purine adducts that interfere with replication.

Y-family DNA polymerases have the ability to replicate through DNA lesions with remarkably diverse strategies⁶. Pol η , for example, can replicate through an ultraviolet-induced *cis-syn* thymine–thymine (T–T) dimer efficiently and accurately^{7–12} and, correspondingly, mutations in Pol η in humans cause a cancer-prone syndrome, the variant form of xeroderma pigmentosum^{11,12}. Unlike Pol η , however, Pol ϵ is unable to replicate through a *cis-syn* T–T dimer¹. Crystal structures of Y-family DNA polymerases include those of yeast Pol η and archaeal Dbh and Dpo4, with only the last enzyme in ternary complex with a template primer and an incoming nucleotide^{13–16}. Pol η , Dbh and Dpo4 are shaped like a right hand, similar to replicative DNA polymerases¹⁷, with palm, fingers and thumb domains, but they contain an additional ‘polymerase-associated domain’ (PAD), which is also referred to as a ‘little finger and wrist’. To our knowledge, the structure of hPol ϵ presented here is the first human Y-family DNA polymerase solved in ternary complex with a template primer and an incoming nucleotide.

The complex contains two hPol ϵ molecules, one at the replicative end of the template primer and the other at the blunt end (Fig. 1). The two hPol ϵ molecules are related by pseudo two-fold symmetry and have very similar structures (root mean square deviation of 0.56 Å for all C α atoms), despite the fact that the polymerase at the replicative end has DNA continuing all the way up to its active site, whereas the active site of the polymerase at the blunt end is devoid of DNA (Fig. 1). The replicative end of the template primer is held in a ‘triangular-like’ embrace by the hPol ϵ palm, fingers and thumb domains, and the PAD unique to Y-family polymerases. The palm (residues 27–36 and 99–224) contains the active-site residues, Asp 34, Asp 126 and Glu 127, that catalyse the nucleotide-transfer

reaction (Fig. 2a). The thumb (residues 224–289) and the PAD (residues 300–414) are connected by a long linker and approach the DNA from opposite sides, with the thumb grazing the minor groove and the PAD docking into the major groove (Fig. 1). The fingers domain (residues 37–98) drapes over the incipient base pair in the active site formed between the templating A and the incoming dTTP (Fig. 1, 2a).

hPol ϵ straddles the blunt DNA end akin to the replicative end, with the PAD and thumb/palm domains on opposite sides of the DNA axis and the fingers domain on top (Fig. 1). Because there is no incoming nucleotide or templating base in the active site of hPol ϵ at the blunt end, however, the residues of the fingers domain that normally stabilize the incoming nucleotide (Tyr 68, Arg 71 and Lys 214) are oriented differently (Fig. 2a). The interface between the two hPol ϵ molecules is small: about 760 Å² of surface area is buried between the thumb domain of one polymerase and the PAD of the other. The fact that hPol ϵ has affinity for blunt-ended DNA in our complex raises the possibility of a role in other forms of DNA repair, such as double-strand break repair by non-homologous end joining, and might be important for its role in somatic hypermutation¹⁸.

The base-pairing in the active site at the replicative end is Hoogsteen and not Watson–Crick (Fig. 2). Characteristic of Hoogsteen base-pairing, the templating A rotates about its glycosidic bond to the *syn* conformation. Two hydrogen bonds are established between the ‘Hoogsteen edge’ of the template A (N7 and N6) and the Watson–Crick edge of the incoming dTTP (N3 and O4), which remains in *anti* conformation (Fig. 2b). The C1’–C1’

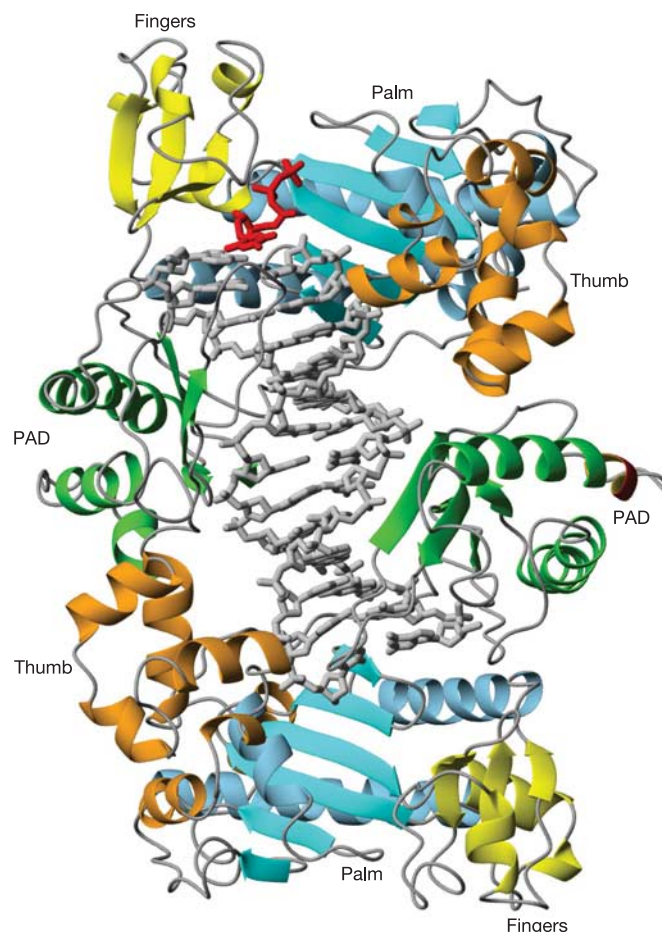


Figure 1 Structure of the hPol ϵ –DNA–dTTP ternary complex. The palm, fingers, thumb domains and the PAD in the two hPol ϵ molecules are shown in blue, yellow, orange and green, respectively. DNA is shown in grey and the incoming dTTP in red.

distance across the incipient A(*syn*)·T(*anti*) base pair is only 8.5 Å, as compared with about 10.5 Å in a canonical Watson–Crick base pair. Hoogsteen base-pairing affords a structural context for the high fidelity and catalytic efficiency of hPol η opposite template A, and its low fidelity and efficiency opposite template T. Opposite template A(*syn*), only the incoming dTTP can make two hydrogen bonds and satisfy the constraint of a short C1'–C1' distance across the base pair. Although incoming dGTP can also make two hydrogen bonds with A(*syn*), the resulting base pair is non-isosteric (C1'–C1' $\approx$ 11.1 Å). Opposite template T(*syn*), there are few hydrogen-bonding opportunities with any incoming nucleotide because T lacks a Hoogsteen edge. Notably, both dCTP and dTTP can make a single hydrogen bond opposite template G(*syn*) at neutral pH¹⁹. Thus, Hoogsteen base-pairing may also account for the intermediate efficiency and fidelity of hPol η opposite G.

What drives the templating A to the *syn* conformation? Relative to the template primer, the fingers domain in hPol η is in a position similar to its position in the Dpo4 ternary complex, but the residues apposed to the templating base are bulkier (Leu 62, Val 64 and Gln

59 in hPol η compared with Ala 42, Ala 44 and Val 32 in Dpo4; Fig. 2a, c). Leu 62, Val 64 and Gln 59 bear down on the templating A, tilting ($\sim 30^\circ$) and rotating it to the *syn* conformation ($\chi = 46^\circ$), with a significant restructuring of the sugar-phosphate backbone. Indeed, if the templating A were to assume the *trans* configuration in hPol η as it does in the Dpo4 complex, Leu 62 would severely clash with it (Fig. 2a, c). The *syn* configuration is stabilized by the sugar of templating A fitting into a small hydrophobic cavity delineated by the aliphatic portions of Gln 59 and Lys 60, and by hydrogen bonds between the phosphate on its 3' side and Lys 60 from the fingers domain and Ser 307 from the PAD. Leu 62 and Lys 60 are unique to hPol η and are the two most essential residues in imposing the *syn* configuration. In human and yeast Pol η , the structural equivalent of Leu 62 (Gly 42 and Ser 58, respectively) is smaller. In addition, superposition of the yeast apo-Pol η structure onto that of hPol η suggests that the fingers domain is farther away from the templating base in Pol η than in hPol η . Thus, unlike Pol η , whose active site seems capable of accommodating a *cis-syn* T–T dimer, as well as incorporating nucleotides that form correct Watson–Crick base pairs,

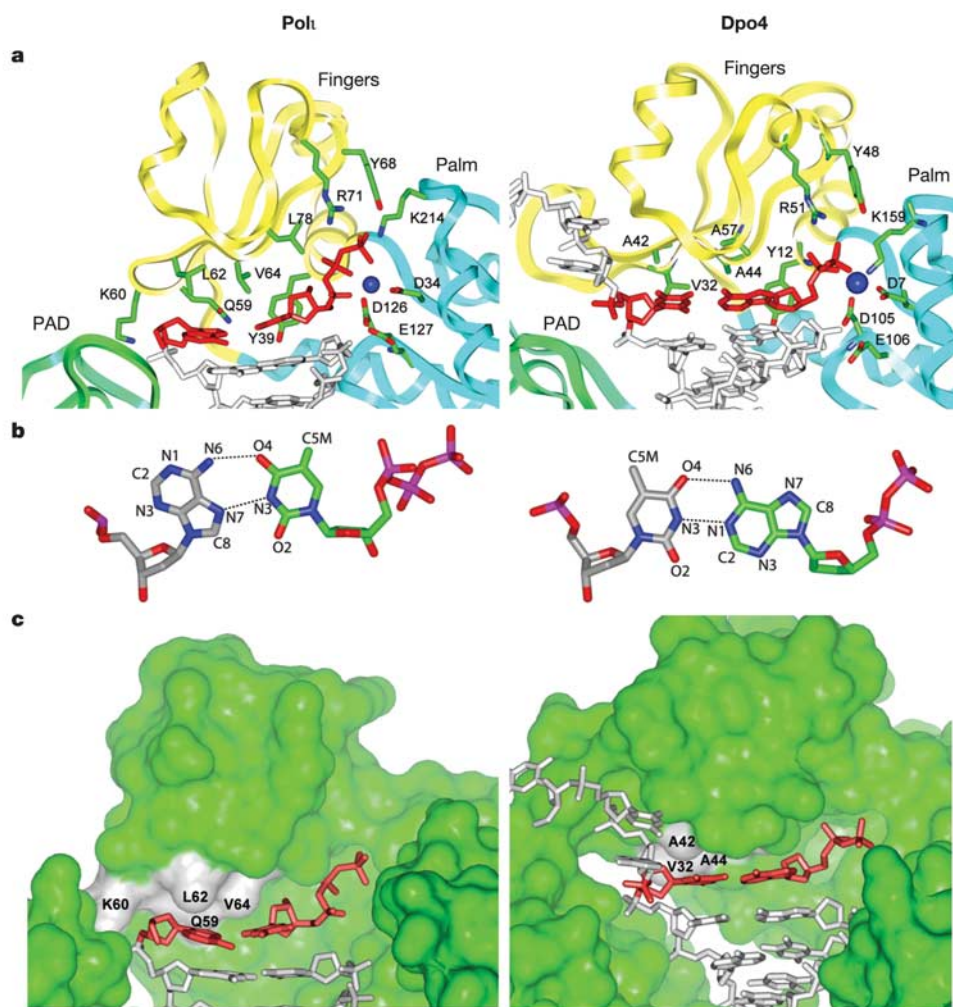


Figure 2 Comparison between hPol η and Dpo4. **a**, Close-up views of the active sites of hPol η and Dpo4. The fingers and palm domains are shown in yellow, blue and green, respectively. The DNA is shown in grey, and the templating nucleotide and the incoming nucleotide (dTTP in hPol η and ddADP in Dpo4) in red. The metal ion is dark blue. Highlighted and labelled are the catalytic residues (Asp 34, Asp 126 and Glu 127 in hPol η and Asp 7, Asp 105 and Glu 106 in Dpo4), the dNTP-bonding residues (Tyr 39, Tyr 68, Arg 71 and Lys 214 in hPol η and Tyr 12, Tyr 48, Arg 51 and Lys 159 in Dpo4), and the

residues close to the templating base (Gln 59, Lys 60, Leu 62 and Val 64 in hPol η and Val 32, Ala 42 and Ala 44 in Dpo4). Note that the templating base is in *syn* conformation in the hPol η active site and in *trans* conformation in the Dpo4 active site. **b**, Close-up views of Hoogsteen base-pairing (dA-dTTP) in the active site of hPol η and Watson–Crick base-pairing (dT-ddADP) in the active site of Dpo4. **c**, Close-up views of the nascent base pair fitting against the hPol η and Dpo4 molecular surfaces. The residues apposed to the templating base are shown in grey and labelled on the molecular surface.

hPol κ cannot accommodate this lesion in its active site.

The dTTP triphosphate moiety zigzags between the fingers and palm domains, making hydrogen bonds with Thr 65, Tyr 68 and Arg 71 from the fingers domain and Lys 214 from the palm domain (Fig. 2a). Tyr 68, Arg 71 and Lys 214 are conserved in all Y-family polymerases, and mutation of the analogous residues in yeast Pol η (Tyr 64, Arg 67 and Lys 159) diminishes the nucleotide incorporation ability of the polymerase²⁰. The dTTP sugar packs against the aromatic ring of Tyr 39, buttressed by a hydrogen bond between its 3'-OH and the main-chain nitrogen of the tyrosine (Fig. 2a). This interaction is compatible with the 'steric gate' mechanism postulated for the exclusion of ribonucleotides by Y-family polymerases, whereby Tyr 39 would sterically discourage nucleotides with a 2'-OH from binding and being incorporated²¹. The catalytic residues, Asp 34, Asp 126 and Glu 127, are clustered between dTTP and the primer terminus. Asp 34 and Asp 126, the main-chain oxygen of

Val 35 and the unesterified oxygens of the α -, β - and γ -phosphates coordinate a single metal ion in the active site (Fig. 2a).

The position of this metal is nearly identical to that of 'metal B' in replicative DNA polymerases¹⁷. For example, as in hPol κ , metal B in *Taq* and T7 DNA polymerases is coordinated by two active-site residues, a main-chain oxygen, and all three phosphates of the incoming nucleotide^{22,23}. Unexpectedly, there is no density for a second metal ion analogous to 'metal A' in replicative polymerases, which is thought to activate the primer 3'-OH. However, Glu 127 is located closer to the primer terminus (~ 3 Å) than is the structurally analogous Glu 655 in T7 DNA polymerase (~ 6 Å), and it may therefore have a more direct role (as a base) in the activation of the primer 3'-OH. In all, the hPol κ active site is well poised for catalysis, with the putative primer 3'-OH aligned for a nucleophilic attack on the dTTP α -phosphate.

Most DNA contacts stem from the PAD (Fig. 3). The PAD has moved by as much as 6 Å and is rotated through 29° from its position in Dpo4 (on the basis of a superposition of their palm domains; Supplementary Fig. 1). The DNA is 'towed' by the PAD; thus, many of the DNA contacts are similar to those in the Dpo4 ternary complex. A key difference is that Arg 343 in hPol κ penetrates the major groove to contact bases directly (Fig. 3), whereas in Dpo4 all of the contacts from the PAD are restricted to the DNA backbone. The PAD makes fewer protein-protein contacts with the fingers domain in hPol κ than in Dpo4, and there is no extended crevice between these two domains in hPol κ (as there is in Dpo4) to accept the unpaired nucleotides on the 5' side of the templating base (Fig. 2a, c). Consequently, these unpaired nucleotides are disordered, and the two ends of the template primer look similar enough to explain in part the affinity of hPol κ to blunt-ended DNA in our complex. The thumb domain is rotated through 33° from its position in Dpo4 and approaches the minor groove at a much more angular orientation (Supplementary Fig. 1). These large movements in the PAD and the thumb domain illustrate the versatility of Y-family polymerases in binding to DNA.

The ternary complex of hPol κ with a template primer and an incoming nucleotide reveals a polymerase specialized for Hoogsteen base-pairing in the active site. The templating base is driven to the *syn* conformation by special features of the hPol κ active site and not by any crystal contacts. Thus, hPol κ is fundamentally different from all other DNA polymerases, which almost all impose Watson-Crick base-pairing between unmodified bases in the active site. Hoogsteen base-pairing offers an elegant basis for the high fidelity and catalytic efficiency of hPol κ opposite template A, and its low fidelity and efficiency opposite template T. In fact, the fidelity opposite template T is so poor that hPol κ inserts an incorrect dGTP roughly 10 times better than it inserts the correct dATP. Exploration of these subtle differences in nucleotide incorporation will require a knowledge of the structures of hPol κ with different template bases and incoming nucleotides (which are likely to be less stably bound).

Hoogsteen base-pairing can also potentially serve as a mechanism to displace adducts that interfere with replication. One could, thus, imagine a role for hPol κ in the bypass of minor-groove adducts of purines, such as those covalently attached to the minor-groove N3 of A or the N² group of G. The rotation of an A or a G from *anti* to *syn* (congruent to Hoogsteen base-pairing) positions N² and N3 in the major groove, where there is much less steric interference. Indeed, hPol κ has been found to promote replication through an N²-adducted G that adopts a *syn* conformation in DNA. Acrolein, an α,β -unsaturated aldehyde, is generated *in vivo* as the end product of lipid peroxidation and during oxidation of polyamines. The reaction of acrolein with the N² of G in DNA, followed by ring closure at N1, leads to formation of the cyclic adduct γ -hydroxy-1,N²-propano-2'-deoxyguanosine (γ -HOPdG), which presents a strong block to synthesis by DNA polymerases. As a templating residue, the ring closed form of γ -HOPdG would adopt a *syn* conformation, presenting its Hoogsteen edge for base-pairing with

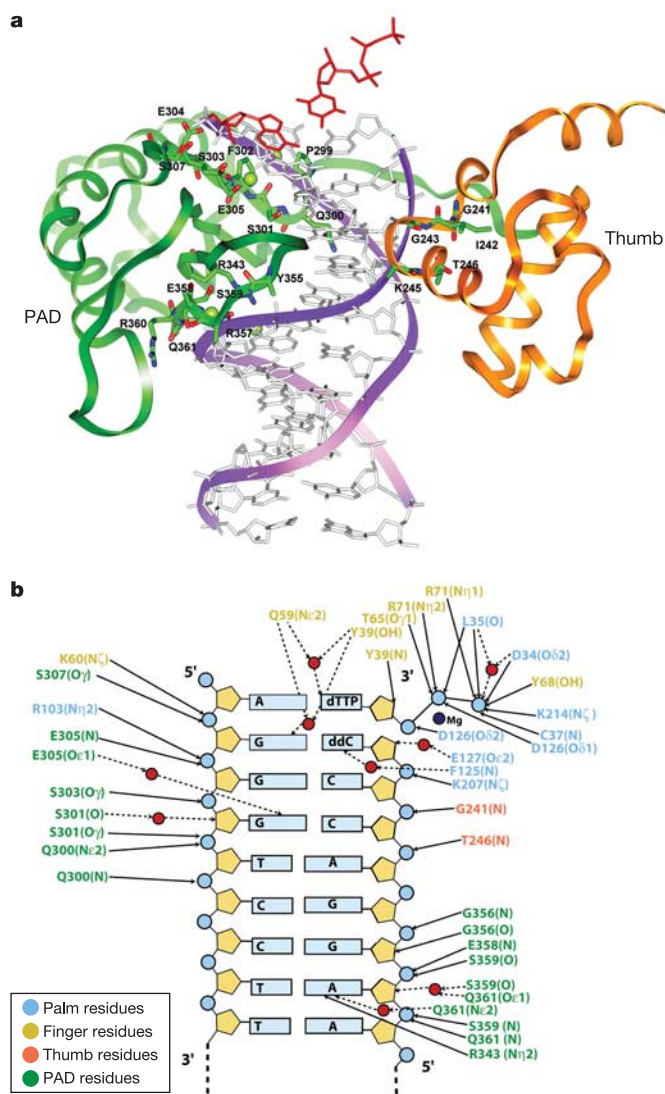


Figure 3 Contacts between DNA and hPol κ . **a**, The hPol κ PAD and the thumb domain grab the DNA from opposite sides. The interacting residues are shown in elemental colours, and the water molecules mediating hydrogen bonding between protein and DNA are shown as yellow-green balls. The side chains of Lys 245, Tyr 355, Arg 357 and Glu 358 are disordered. **b**, Schematic view of the hydrogen bonds between hPol κ and the DNA and incoming dTTP. Hydrogen bonds mediated by water (red spheres) are indicated by broken lines. Note that many of the hydrogen bonds to the DNA are made by hPol κ main-chain atoms, whereas hydrogen bonds to dTTP are primarily made by side chains.

a C. hPol₁ incorporates a C opposite this lesion as efficiently as opposite an undamaged G, thereby providing an effective means for replication through such minor-groove adducts²⁴. □

Methods

Protein and DNA preparation

The glutathione S-transferase (GST) and hPol₁ (residues 1–420) fusion protein was expressed in yeast strain BJ5464 from plasmid pBJ941, and then purified and cleaved as described for yeast Pol_η¹³. To prepare selenomethionine (SeMet)-labelled hPol₁, we expressed plasmid pBJ1066 in *Escherichia coli* strain B834, which is auxotrophic for methionine, and grew the cells in M9 minimal medium. The primer was synthesized as 12- and 13-nucleotide oligonucleotides, with the latter containing a dideoxycytosine at its 3' end (5'-GGGGGAAGGACCC^{4d}-3'). These were annealed with an 18-nucleotide template (5'-TTCTAGGGTCCTTCCCC-3') to yield 12/18 and 13/18 primer templates. The template was also synthesized with three thymines substituted by 5-bromouracil (5'-TTCAAGGGU^BCCU^BU^BCCCC-3'), and then annealed with the 13-nucleotide primer to yield a 13/18(Br) primer template.

Cocrystallization

We obtained three different cocrystals: SeMet, Native1 and Native2. SeMet cocrystals were obtained by incubating SeMet hPol₁ with the 12/18 primer template in a molar ratio of 1:1.2, plus 5 mM dTTP and 5 mM MgCl₂. Native1 cocrystals were obtained by mixing native hPol₁ with the 13/18 primer template plus 5 mM dTTP and 5 mM MgCl₂. The Native2 cocrystals were obtained by incubating native hPol₁ with the 13/18(Br) primer template plus 15 mM dTTP and 5 mM MgCl₂. For each, the complex was crystallized from solutions containing 10–15% PEG 5000 MME and 0.2–0.4 M (NH₄)₂SO₄ in 0.1 M MES buffer (pH 6.5). The three cocrystals belong to the space group P6₃, with identical cell dimensions of *a* = 98.5 Å, *b* = 98.5 Å, *c* = 203.7 Å and $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$.

Structure determination

Multiwavelength anomalous diffraction (MAD) data on cryo-cooled SeMet cocrystals (2.6 Å) were measured at the Advanced Photon Source (APS, beamline 14-ID) at three wavelengths, corresponding to the edge and peak of the Se K edge absorption profile plus a remote point (Supplementary Table 1). The Native1 and Native2 data were also measured at the APS at beamlines 14-ID (2.1 Å) and 17-ID (2.3 Å), respectively. We found the positions of nine selenium atoms from the SeMet data using the program SOLVE²⁵. The initial experimental phases (2.8 Å) from these selenium positions were applied to Native1 data, and then extended to 2.1 Å with solvent flattening by the program CNS²⁶. This yielded a readily interpretable electron density map that was used to build the initial complex. Unexpectedly, two hPol₁ molecules were bound to the DNA, related by a pseudo-dyad axis perpendicular to the DNA axis. The two hPol₁ molecules were built independently and refined, but there was no density for the incoming nucleotide.

The Native1 structure was used as a molecular replacement model to phase the Native2 complex (crystallized with brominated DNA and dTTP instead of dTTTP) with the program AmoRe²⁷. The Native2 structure showed clear electron density for the incoming dTTP. Anomalous difference Fourier maps revealed six bromine peaks, suggesting that the complex packed in two orientations in the crystal. The Native2 structure was refined in two orientations with CNS, with the bromine atoms helping to fix the register of the DNA. After iterative rounds of refinement, model building with program O²⁸ and water picking, the *R*_{free} dropped to 28.6%. The final Native2 model includes residues 27–414 for molecules A and B, DNA (electron density for four of the five unpaired template residues towards the 5' end was not visible), dTTP and 438 water molecules. The model has good stereochemistry: 82% of the residues are in the most favoured regions of a Ramachandran plot and only 0.5% are in the disallowed regions.

Received 2 March; accepted 28 May 2004; doi:10.1038/nature02692.

- Johnson, R. E., Washington, M. T., Haracska, L., Prakash, S. & Prakash, L. Eukaryotic polymerases ϵ and ζ act sequentially to bypass DNA lesions. *Nature* **406**, 1015–1019 (2000).
- Tissier, A., McDonald, J. P., Frank, E. G. & Woodgate, R. Pol_η, a remarkably error-prone human DNA polymerase. *Genes Dev.* **14**, 1642–1650 (2000).
- Zhang, Y., Yuan, F., Wu, X. & Wang, Z. Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase ϵ . *Mol. Cell. Biol.* **20**, 7009–7108 (2000).
- Haracska, L. *et al.* Targeting of human DNA polymerase ϵ to the replication machinery via interaction with PCNA. *Proc. Natl Acad. Sci. USA* **98**, 14256–14261 (2001).
- Washington, M. T., Johnson, R. E., Prakash, L. & Prakash, S. Human DNA polymerase ϵ utilizes different nucleotide incorporation mechanisms dependent upon the template base. *Mol. Cell. Biol.* **24**, 936–943 (2004).
- Prakash, S. & Prakash, L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev.* **16**, 1872–1883 (2002).
- Johnson, R. E., Prakash, S. & Prakash, L. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase Pol_η. *Science* **283**, 1001–1004 (1999).
- Washington, M. T., Johnson, R. E., Prakash, S. & Prakash, L. Accuracy of thymine-thymine dimer bypass by *Saccharomyces cerevisiae* DNA polymerase η . *Proc. Natl Acad. Sci. USA* **97**, 3094–3099 (2000).
- Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. Fidelity of human DNA polymerase η . *J. Biol. Chem.* **275**, 7447–7450 (2000).
- Washington, M. T., Prakash, L. & Prakash, S. Mechanism of nucleotide incorporation opposite a thymine-thymine dimer by yeast DNA polymerase η . *Proc. Natl Acad. Sci. USA* **100**, 12093–12098 (2003).

- Johnson, R. E., Kondratieff, C. M., Prakash, S. & Prakash, L. *hRAD30* mutations in the variant form of xeroderma pigmentosum. *Science* **285**, 263–265 (1999).
- Masutani, C. *et al.* The *XPV* (xeroderma pigmentosum variant) gene encodes human DNA polymerase η . *Nature* **399**, 700–704 (1999).
- Trincao, J. *et al.* Structure of the catalytic core of *S. cerevisiae* DNA polymerase η : implications for translesion DNA synthesis. *Mol. Cell* **8**, 417–426 (2001).
- Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102 (2001).
- Zhou, B.-L., Pata, J. D. & Steitz, T. A. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol. Cell* **8**, 427–437 (2001).
- Silvian, L. E., Toth, E. A., Pham, P., Goodman, M. F. & Ellenberger, T. Crystal structure of a DinB family error-prone DNA polymerase from *Sulfolobus solfataricus*. *Nature Struct. Biol.* **8**, 984–989 (2001).
- Steitz, T. A. DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem.* **274**, 17395–17398 (1999).
- Faill, A. *et al.* Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase ι . *Nature* **419**, 944–947 (2002).
- Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson–Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**, 3497–3531 (2002).
- Johnson, R. E., Trincao, J., Aggarwal, A. K., Prakash, S. & Prakash, L. Deoxynucleotide triphosphate binding mode conserved in Y family DNA polymerases. *Mol. Cell. Biol.* **23**, 3008–3012 (2003).
- DeLucia, A. M., Grindley, N. D. & Joyce, C. M. An error-prone family Y DNA polymerase (DinB homolog from *Sulfolobus solfataricus*) uses a 'steric gate' residue for discriminating against ribonucleotides. *Nucleic Acids Res.* **31**, 4129–4137 (2003).
- Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 (1998).
- Li, Y., Korolev, S. & Waksman, G. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J.* **17**, 7514–7525 (1998).
- Washington, M. T. *et al.* Efficient and error-free replication past a minor groove DNA adduct by the sequential action of human DNA polymerases ϵ and κ . *Mol. Cell. Biol.* **24**, 5687–5693 (2004).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Brunker, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank staff at the APS for facilitating data collection; and T. Edwards and C. Escalante for help with structure determination. This work was supported by grants from the NIH (to A.K.A. and L.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.K.A. (aggarwal@inka.mssm.edu). Coordinates have been deposited in the Protein Data Bank under accession code 1T3N.

erratum

Evidence for dynamically organized modularity in the yeast protein–protein interaction network

Jing-Dong J. Han, Nicolas Bertin, Tong Hao, Debra S. Goldberg, Gabriel F. Beriz, Lan V. Zhang, Denis Dupuy, Albertha J. M. Walhout, Michael E. Cusick, Frederick P. Roth & Marc Vidal

Nature **430**, 88–93 (2004).

This Letter (doi:10.1038/nature02555) was published ahead of print and should have included the line ‘Published online 9 June 2004’. □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequinot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: r.asami@naturejp.com

naturejobs

Back on track?

Most researchers like the idea of tenure — especially once they have it. But many others, particularly those pursuing it, are not so thrilled about the way in which it is awarded. Indeed, a system primarily based on publications and citations seems a bit antiquated. It doesn't necessarily reflect the way science is now being done, nor does it reward some professional practices most reasonable people would agree are positive. But a few simple reforms could improve the system.

One of the biggest problems with the current system is that it treats scientists as isolated geniuses, locked in a lab. Intentional or not, this is the basic effect of rewarding first authorship. But science is now more often a collaborative effort, and credit should be spread more equally among participants. It doesn't make sense that someone doing data collection at a telescope or coordinating high-throughput microarray analysis should be left behind just because they seldom have the opportunity to write up the breakthroughs they help to facilitate.

And most scientists agree that some activities beyond publication are also worthwhile. But why should they take the time to mentor thoroughly, teach well or communicate science to the public if they aren't promoted for it?

It would be nice to see such salutary activity quantified, then rewarded. Perhaps there should be a system that rewards mentoring, maybe with bonus points for those who mention career options outside academia, aid networking and have a good track record for finding jobs for their protégés. And as for team players, especially in interdisciplinary projects, how about taking into account how many papers they have made possible, even though they may not have done the primary analysis? Facilitating other people's scientific discoveries and career progress is an important part of the scientific world and should be recognized.

Paul Smaglik

Naturejobs editor



Contents

CAREERS AND RECRUITMENT

The growing world of
developmental biology p382

CAREER VIEW

Bricks & Mortar

San Jose BioScience Incubator
and Innovation Center

Graduate Journal

Time bends

Movers

Daniel Carucci

p384

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Developing themes

Matching the march of evolution, developmental biology is branching out to encompass a wide variety of disciplines. As a result, recruiters want skills as well as qualifications, says Ricki Lewis.

African clawed frogs, fruitflies, sea urchins, chickens, flour beetles and cockroaches may sound like a motley crew, but they all helped to give rise to experimental developmental biology in the twentieth century. They provided a parade of diversity that under scrutiny revealed the similarities of early animal development. In the 1980s, recombinant-DNA technology enabled investigators to isolate the contributions of individual genes and to tie experimental embryology, developmental genetics and molecular biology together. With the dawn of this century, genomics and bioinformatics have redefined the field, revealing global patterns of gene expression within the individual organism and across evolutionary time.

Although new technologies have allowed researchers to dissect the controls of developmental pathways, a broader view has embraced evolution and ecology. The overlap of areas that defines developmental biology in its latest guise makes quantifying career options difficult, but principal investigators report that flexibility works to a fledgling researcher's advantage — skills, rather than a specific degree, are being sought after.

"There are very few rules in this area, as job descriptions can vary widely," says Ramesh Shivdasani, associate professor of medicine at Harvard Medical School in Boston, Massachusetts. "In general, trainees with a background in developmental biology are considerably more attractive as recruits today than a decade ago, when the field was more descriptive."

As in evolution itself, successful developmental biologists have adapted to change, while appreciating the legacy of the field. "I am always looking for people who have a combination of experience in molecular biology and a strong understanding of embryology," says Stephen Duncan, associate professor of cell biology, neurobiology and anatomy at the Medical College of Wisconsin in Milwaukee. "Surprisingly, people with skills in both are actually quite rare."

And training on any species is valuable. "A scientist working with flies or worms will be able to do research in a fertility lab or a biotech company," says Randy Jirtle, whose lab investigates genomic imprinting at Duke University Medical Center in Durham, North Carolina. The joys of switching between species and from single gene to entire organism are reflected in career options. Academia

remains the stronghold of job opportunities, but the new directions have led to eclectic applications, from aquaculture to zoo management. Opportunities also include traditional academic research, industrial science, teaching, and clinical practice and research in reproductive medicine.

Although options abound beyond universities, the developmental-biology jobs pendulum seems to have swung back to academia, says Jerry Shay, a professor of cell biology and neuroscience at the University of Texas Southwestern Medical Center in Dallas. About ten years ago, half his students and postdocs went on to industry. Now about two-thirds opt to stay in academia. And at Indiana University, only a "tiny percentage" of graduate students in molecular, cell and developmental biology enter industry, says Gretchen Clearwater, adviser for graduate affairs.

The picture is a bit different at the University of California, Irvine, says Peter Bryant, director of the graduate programme in molecular biology, genetics and biochemistry there. About half of his 53 trainees have gone on to non-traditional careers, including art, optometry, writing, teaching, toxicology, epidemiology and even directing educational programmes at a zoo.

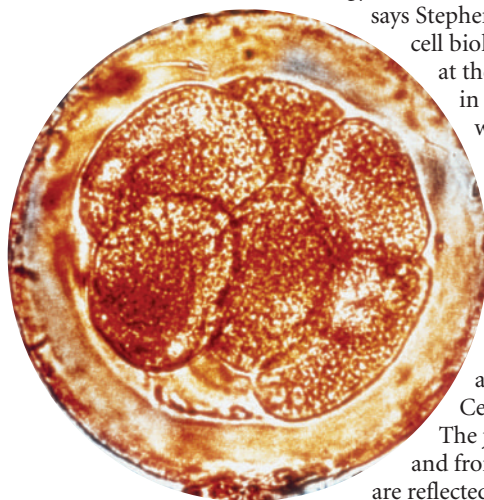
BROAD SPECTRUM

The job spectrum in academia has also broadened. For example, several new academic positions seek expertise in evolutionary developmental biology, better known as 'evo-devo'. In evo-devo, "the most important skills are fluency in developmental genetics, molecular genetics and evolutionary biology, not only in terms of techniques and approaches, but in types of questions asked," says Vivian Irish, director of undergraduate studies in biology at Yale University in New Haven, Connecticut.

Genomics is another focus that has profoundly altered how academic developmental biologists frame questions, says Shivdasani. "An ability to harness genomic tools is a significant asset in any potential recruit's skill set," he adds. The genomic view, ironically, takes the field full circle — from whole organism, to single genes and now to comparing species.

Stem-cell biology, still in its infancy, is also part of developmental biology. It touches academia, industry and the clinic. "Stem-cell biology and tissue engineering link developmental biology even better with medicine than before," says Rolf Zeller, professor of developmental genetics at the University of Basel Medical School in Switzerland. Even restrictions in some nations, such as the United States, on the use of human embryonic stem cells cannot slow the momentum. "Stem cells rock," says Shay, "and will be here for a long time."

Heart of the matter: watching animals develop is what gives many researchers satisfaction.



EDLMANN/SPL



EIDELMANN/SPL

Growing up: developmental biology has expanded to encompass fertility treatment.

Another area of research that provides a perfect match for today's developmental biologist is working with transgenic organisms. "We are developing transgenic chickens to express proteins of therapeutic interest in their egg white," says Leandro Christmann, director of transgenic technology at AviGenics in Athens, Georgia. This involves transplanting unspecialized cells into embryos and getting them to differentiate and produce the desired protein. "One needs to understand the principles of developmental biology to accomplish this," he adds.

In industry, 'hands-on' ability — manipulating gametes and embryos — is crucial. Positions at biotech giant Serono International in Geneva, for example, often accept any life-sciences degree, but require specific skills such as micromanipulation, embryo transfer or running stem-cell cultures. "We take people with many different academic backgrounds. We see developmental biology as a training in basic biology, as well as in key processes in disease such as differentiation and apoptosis," says Tim Wells, Serono's senior executive vice-president for research. Preparation for a job in industry may require more than one postdoctoral position to achieve adequate breadth, adds Mark Pittenger, vice-president for research at Osiris Therapeutics, a regenerative-medicine company in Baltimore, Maryland.

CLINICAL TRAINING

A clinical career, like a biotech/industrial track, also includes advanced training. For example, in the United States, working with infertility patients requires certification by the American Board of Bioanalysis, which entails holding a degree, having relevant experience, performing procedures and passing an

exam. The American Association of Tissue Banks certifies reproductive-cryotechnology specialists, who work with frozen sperm and embryos.

Those who wish to abandon worms or flies for humans may benefit from a programme such as the postdoctoral opportunities at the Center for Research on Reproduction and Women's Health at the University of Pennsylvania in Philadelphia. "We focus on gamete biology, so the type of research training our people get is very relevant to clinical embryology," says director Jerome Strauss.

THE TEACHING TRACK

Developmental biology attracts teachers: several researchers mention graduate students who entered their labs with precisely this goal, rather than pursuing teaching when other options failed.

Sam Rhine, founder of the Genetic Education Center in Indiana, turned his love of teaching into annual genetic-update conferences for teachers and students. Rhine, who has done his road show full-time since 1994, visits high schools, educational conferences and colleges wanting to entice better biology students. "I have about 350 of the best students in the United States in my audiences daily. What teacher wouldn't love an opportunity like that?" he says.

Whether a developmental biologist works at a university, agribiotech company or fertility clinic, all begin with a shared wonder at the unfolding of a new organism from an initial cell. "For most of us, the love of looking at organisms develop is what got us hooked," says Barbara Wakimoto, a professor of zoology at the University of Washington, Seattle. And even in this ever-changing field, that fascination is timeless. ■

Ricki Lewis is a science writer based in Scotia, New York.

GRADUATE JOURNAL

Time bends

I'm not sure how it works, but I think I've experienced time speeding up, slowing down and stopping during these graduate school years.

I now expect time to accelerate, and more computers to crash, as a deadline looms. When I am anticipating an exciting result and have to wait for a colorimetry reaction to develop, an hour feels like half a day. An afternoon of dilutions can feel like an eternity of boredom.

Our lab has no windows. It is easy, especially when one works through lunch, to forget what time it is. The lab sits at 43° N; in the depths of winter one can arrive at work under a starry sky, and leave at the end of the day long after the Sun has set on a cold and snowy Toronto. It is spookily easy to lose track of the days passing.

My favourite time-bending experience usually catches me by surprise. When I dive into a problem, a book or an editing project, hours pass unnoticed. It is usually the rumbling of an unattended stomach that finally yanks me back to the reality that I'm late to meet my friends for dinner. I wish I could explain how my tardiness can stem from time-warping in the lab, yet I simultaneously give thanks that this is one effect that I don't need to explain to pass my thesis defence. ■

Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

BRICKS & MORTAR

San Jose BioScience Incubator and Innovation Center

Conventionally, biotechnology incubators provide space for fledgling companies to grow. But a new generation of facilities is springing up that offers additional services to help nurture budding businesses. One such newcomer is the San Jose BioScience Incubator and Innovation Center, which opened last month and is expecting its first tenant in August. The centre is the latest in a number of government-funded facilities that provide a support structure as well as lab space.

The building's 3,400 square metres should comfortably house about 25 companies, which will be able to use the centre's support staff to help establish and meet their business goals. Melinda Richter, managing director of Global Access to

Innovation Networks (GAIN), the Bay Area company that will manage the facilities at the centre, sees the set-up as an "ecosystem".

The companies that secure space in the incubator must first undergo a business assessment with GAIN. This looks at the nascent firm's strengths and weaknesses — including intellectual property, finances and management — and identifies its three most important needs. These are then used as benchmarks, and the firm must meet with GAIN quarterly to check on its progress.

To help the tenant meet the goals, the centre assigns each firm a 'chief-executive coach' from a non-competing company. In addition, the incubator offers access to a database of potential mentors that can help the company to find people such as intellectual property lawyers or contacts at

potential commercial partners. The centre also plans to house several companies 'virtually', by providing them with its services but not space.

The city of San Jose spent \$6.5 million to renovate, equip and manage the centre in order to attract jobs. An earlier incubator, which opened its doors in 1994, focused on information technology and has already attracted \$475 million in venture capital and created 2,500 jobs.

Richter hopes that the combination of service and space at the biosciences centre will serve as "an international landing pad" for companies from abroad looking to get a toehold in the United States. She is optimistic about the centre's future — after all, the Milken Institute recently listed San Jose as one of the top ten sustainable biotech cities. ■

Paul Smaglik is editor of Naturejobs.
♦ www.wtc-sf.org/biosanJose.html

MOVERS

Daniel Carucci, director, Grand Challenges in Global Health, National Institutes of Health



Daniel Carucci joined the US Navy to finance his medical education. But during his service, he found that, rather than an obligation, his military stint opened up opportunities he never would have considered had he not joined.

Carucci learned to fly fighter jets, became interested in tropical medicine and got involved in genomics. "When opportunities presented themselves to

me, I often took them, even if they were not on my critical path," Carucci says. This month, his path hit a critical junction as he retired from the Navy after 20 years — 17 more than were needed to pay for his medical training — to take on a position tackling global health challenges.

His military stint stimulated his interest in infectious diseases. After training as a flight surgeon, Carucci was deployed for months at a time in Asia, Central America and Africa. On breaks from caring for pilots and their families, Carucci would visit local hospitals and clinics, where he encountered tropical medicine in the tropics.

After he won an award for flight surgeon of the year in 1989, the Navy allowed him to take a master's degree at the London School of Hygiene and Tropical Medicine, where he got his first taste of hands-on research. His appetite whetted, he went on to do a PhD.

After Carucci finished this, Steve Hoffman, then director of the malaria programme at the Navy Medical Research Center in Silver Spring, Maryland, recruited Carucci to the Navy's malaria-vaccine programme. Carucci helped lead the malaria sequencing effort and got interested in proteomics and functional genomics. When Hoffman left for the genomics company Celera in 2001, Carucci took over his post, and branched out into immunology and vaccine research. Both positions taught him how to pick up new scientific fields quickly and to lead large teams with skills ranging from basic biology to regulatory affairs.

In his new job, based in Bethesda, Maryland, the remit will be even broader. The National Institutes of Health's Grand Challenges in Global Health programme will extend well beyond malaria. The job will give Carucci plenty of opportunities to follow his curiosity. ■

CV

2001–04: Director, malaria programme, Naval Medical Research Center, Silver Spring, Maryland

1999–2000: Director, genomics and bioinformatics, malaria programme, Naval Medical Research Center, Silver Spring, Maryland

1995–99: Research physician/molecular biologist, malaria programme, Naval Medical Research Center, Silver Spring, Maryland

1995: PhD molecular biology, London School of Hygiene and Tropical Medicine